UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Reconhecimento Automático de Expressões Faciais Baseado em Características Geométricas

Jovan de Andrade Fernandes Junior

SÃO CRISTÓVÃO/ SE
2016

UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Jovan de Andrade Fernandes Junior

Reconhecimento Automático de Expressões Faciais Baseado em Características Geométricas

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação (PROCC) da Universidade Federal de Sergipe (UFS), como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. D.Sc. Leonardo Nogueira Matos

SÃO CRISTÓVÃO/ SE
2016

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL UNIVERSIDADE FEDERAL DE SERGIPE

Fernandes Junior, Jovan de Andrade
F363r Reconhecimento automático de expressões faciais baseado em características geométricas / Jovan de Andrade Fernandes Junior ; orientador Leonardo Nogueira Matos. – São Cristóvão, 2016.
55 f. : il.

Dissertação (mestrado em Ciências da computação)– Universidade Federal de Sergipe, 2017.

1. Programas de computador. 2. Software - Desenvolvimento. 3. Tecnologia da informação. 4. Comunicação da tecnologia. I. Matos, Leonardo Nogueira. II. Título.

CDU: 004.4

Jovan de Andrade Fernandes Junior

Reconhecimento Automático de Expressões Faciais Baseado em Características Geométricas

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação (PROCC) da Universidade Federal de Sergipe (UFS) como parte de requisito para obtenção do título de Mestre em Ciência da Computação.

BANCA EXAMINADORA

Prof. D.Sc. Leonardo Nogueira Matos, Presidente
Universidade Federal de Sergipe (UFS)

Prof. Dr. Paulo Salgado Gomes de Mattos Neto, Membro
Universidade Federal de Pernambuco (UFPE)

Prof. Dr. Jugurta Rosa Montalvão Filho, Membro
Universidade Federal de Sergipe (UFS)

UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
NÚCLEO DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Relatório de defesa pública do(a) Senhor(a) **JOVAN DE ANDRADE FERNANDES JUNIOR** no Programa de Ciência da Computação (PROCC) da UFS.

Aos 21 dias do mês de dezembro de 2016, realizou-se a **Defesa de Mestrado** do trabalho intitulado **"Reconhecimento Automático de Expressões Faciais Baseado na Geometria Facial"** sob orientação do Prof. Dr. **Leonardo Nogueira Matos.**

Depois de declarada aberta a sessão, o Presidente da Banca passou inicialmente a palavra ao candidato para exposição e a seguir aos examinadores para as devidas arguições que se desenvolveram nos termos regimentais. Em seguida, a comissão julgadora proclamou o resultado:

Nome Banca Examinadora	Instituição	Assinatura
Leonardo Nogueira Matos	UFS	
Jugurta Rosa Montalvão Filho	UFS	
Paulo Salgado Gomes de Mattos Neto	UFPE	

Dessa maneira o Resultado Final é: X **APROVADO ou** **Reprovado**

Parecer da Banca Examinadora *

- Obs: Se o candidato for reprovado, o preenchimento do parecer é obrigatório

São Cristóvão/SE

Assinatura do Orientador: _____ Assinatura do Aluno: _____

Agradecimentos

Agradeço, em primeiro lugar, a meus pais, Mary Ane e Jovan, pela educação, valores, paciência, amor e dedicação. As minhas irmãs, Katiuska e Jéssika que sempre me apoiaram e aconselharam em todas as decisões que tomei na vida. A minha namorada, Priscilla, que me apoiou nos momentos difíceis de toda a trajetória do programa de mestrado; sem ela certamente não teria tido determinação para finalizar esta etapa na minha vida.

Um trabalho de pesquisa leva tempo, foco, paciência e determinação. Portanto, não poderia deixar de agradecer ao meu orientador, Leonardo, pelo companheirismo, paciência e dedicação exemplar, por sempre possuir algo enriquecedor a acrescentar durante os debates da pesquisa, e por oferecer sua orientação mesmo nos horários em que poderia estar descansando ao lado de sua família.

Aos professores do PROCC, que mudaram a minha maneira de pensar e enxergar o mundo, em especial os professores Jugurta Montalvão, Carlos Estombelo e Hendrik, que foram fundamentais durante esta caminhada.

Aos colegas de trabalho do TRE-SE e da Infox, que gentilmente flexibilizaram meus horários de trabalho para poder frequentar as aulas e reuniões de orientação.

Resumo

Nos últimos anos temos observado grandes avanços na área de Visão Computacional que possibilitaram uma mudança na maneira como nos relacionamos com a máquina. Para alcançar uma efetiva Interface Humano-Computador Inteligente (IHC), além dos movimentos corporais ou comandos vocais, é necessário que a máquina seja capaz de compreender também as expressões faciais dos seres humanos.

Diversos autores buscaram reconhecer expressões faciais mas essa tarefa ainda não é executada com a mesma eficiência que um humano. Este trabalho se utilizou da geometria facial humana para propor dois métodos de seleção de características para reconhecer expressões faciais humanas. O primeiro, intitulado método das Distâncias Empíricas, obteve 77.66% de acurácia, enquanto que o segundo, intitulado método das Distâncias CFS, obteve uma acurácia de 91.33%. Os resultados obtidos foram compatíveis com o atual estado da arte da área de pesquisa.

Abstract

In recent years we have seen great advances in Computer Vision research area that have made possible change the we interact with machines. To achieve an effective Intelligent Human-Computer Interface (IHC), in addition to recognize body movements or vocal commands, it is necessary the machine be able to understand human facial expressions.

Although there are several publications that aims to recognize facial expressions, this task is not yet performed by a machine with the same efficiency as the human being. This work proposes two geometric-based feature selection approaches for facial expression recognition. The first, called Empirical Distances method obtained 77.66% of recognition rate. The second, called CFS Distances method, obtained 91.33% of recognition rate. The results obtained are compatible with the state of the art in this research area.

Lista de Figuras

1.1 Emoções básicas sugeridas por Ekman [1] demonstradas com imagens da base de dados Cohn-Kanade [2] 2

3.1 Visão geral do método proposto. 11
3.2 Conjunto de pontos utilizados para representar os componentes faciais. . . . 13
3.3 Uma forma s pode expressada através de uma forma base s_0 acrescida de uma combinação linear das variações da sua foram s_i. 13
3.4 Modelo linear da forma de um AAM. 15
3.5 Exemplo de instanciação de um AAM. Na parte superior da imagem tem-se o modelo de aparência. Na parte inferior, o modelo da forma. No centro, a concatenação dos modelos gerando uma instância denotada por $M(T(x, p))$ que correspondente a imagem real $I(x)$. 16
3.6 Representação do modelo de formas e o modelo de fragmentos de um CLM. 18
3.7 Visão geral do funcionamento de um CLM. 19
3.8 Exemplos de imagens utilizadas em uma pesquisa experimental realizada com o objetivo de identificar quais as distâncias mais relevantes segundo a visão humana para a classificação de expressões faciais. 22
3.9 Distâncias empíricas consideradas. 23
3.10 Distâncias selecionadas pela seleção de características CFS. 25
3.11 O hiperplano ótimo de separação com margem máxima ρ. Os vetores de suporte são as amostras que satisfazem as equações $g(\mathbf{x}) = 1$ ou $g(\mathbf{x}) = -1$. Adaptado de Hammel [3]. 26
3.12 Mapeamento de dados para um espaço de características de maior dimensão. 26

4.1 Uma possível dobra para o protocolo de amostragem utilizado. 31

A.1 Exemplos de marcação dos prontos de controle em localização equivalente utilizando imagens da base de dados MUCT [4]. 50
A.2 Demonstração das variações de translação, rotação e escala. Adaptado de Baggio [5]. 51
A.3 Resultado da *Análise de Procrustes* 53
A.4 Modos de Variação considerando o objeto face humana. Adaptado de Baggio [5]. 54
A.5 Modelo de perfil de intensidade. Na imagem, cada ponto de referência está representado pelo seu perfil de intensidade médio. 57
A.6 Resultados antes e depois da realização da correspondência com o formato da face. Adaptado de Baggio [5]. 59
A.7 Resultados obtidos com diferentes posicionamentos iniciais do modelo. Adaptado de Cootes *et al.*[6]. 60
A.8 Níveis de resolução. Adaptado de Cootes *et al.* [6]. 60

Lista de Tabelas

1.1 Expressões Faciais Básicas [1] 3

3.1 Distâncias Empíricas utilizadas. 23

4.1 Acurácia média e desvio padrão por classe obtidos com a estratégia de seleção de atributos das *Distâncias Empíricas* aplicada ao problema de 6 classes. 32

4.2 Acurácia média e desvio padrão por classe obtida adicionando a expressão Neutra ao reconhecimento da expressões utilizando a abordagem das *Distâncias Empíricas*. 32

4.3 Acurácia média por classe obtida com a seleção de características das *Distâncias CFS* no problema de 6 classes. 33

4.4 Acurácia média por classe obtida com a seleção de características das *Distâncias CFS* no problema de 7 classes. 34

4.5 Comparação de performance das duas abordagens propostas nesse trabalho utilizando o PDM independente de indivíduo com métodos de reconhecimento de expressões faciais que utilizaram a Cohn-kanade [2] nos experimentos. 36

Sumário

1 Introdução **1**

1.1 Motivação 1

1.2 Expressões Faciais 1

1.3 Objetivos 4

1.4 Organização do Documento 5

2 Revisão Bibliográfica **6**

2.1 Métodos de Extração e Seleção de Características Faciais 7

2.2 Trabalhos Relacionados 8

3 Sistema de Reconhecimento de Expressões Faciais **11**

3.1 Modelos Pontuais de Distribuição 12

3.1.1 Modelos de Aparência Ativa 14

3.1.2 Modelos de Restrições Locais 17

3.2 Extração e Seleção de Características 20

3.2.1 Distâncias Empíricas 21

3.2.2 Correlation Features Selection - CFS 23

3.3 Classificação de Padrões 25

3.3.1 Máquina de Vetores de Suporte 25

3.3.2 Seleção de Parâmetro de Kernel 27

4 Avaliação Experimental **29**

4.1 Condução dos Experimentos 29

4.2 Experimentos 31

4.3 Comparação e Discussão 34

5 Conclusão **37**

Referências **38**

A PDM - Modelos Pontuais de Distribuição **47**

A.1 Histórico 47

A.2 Construindo um Modelo Pontual da Forma do Objeto 49

A.2.1 Posicionamento do pontos de Referência 49

A.2.2 Alinhamento da Base de Treinamento 50

A.2.3 Estudo das Variações Admissíveis 52

A.2.4 Escolha dos Modos de Variação Admissíveis 54

A.3 Utilizando os PDMs em Problemas de Busca em Imagens 55

A.3.1 ASM - Modelo de Forma Ativa 55

Capítulo 1

Introdução

A interação Humano-Computador é uma matéria multidisciplinar que envolve áreas da engenharia, psicologia linguística, dentre outras. Para alcançar uma efetiva Interface Humano-Computador Inteligente (IHCI) é necessário que a máquina seja capaz de interagir naturalmente com o usuário, similar à interação entre humanos.

1.1 Motivação

Durante uma comunicação interpessoal, a informação pode ser transmitida através de diversos meios como auditivo, visual e o tátil. Dentre eles, o visual é o meio que carrega a maior quantidade de informação envolvendo gestos corporais, expressões faciais ou uma combinação destes. Sabendo que os seres humanos são criaturas emocionais cujo estado de espírito pode ser observado em grande parte através de expressões faciais, reconhecer e interpretar expressões faciais humanas torna-se uma tarefa essencial para alcançar uma interação satisfatória entre o homem e a máquina.

1.2 Expressões Faciais

As emoções humanas começaram a ser investigadas por Charles Darwin no seu livro "*The expression of the emotions in man and animals*"[7], publicado em 1872. Após este passo inicial, diversos pesquisadores estudaram o comportamento afetivo humano e, apesar de não ser um consenso no meio científico, vários autores chegaram a conclusão de que existe um

conjunto de emoções que aparentam ser reconhecidas universalmente, mesmo em diferentes culturas e etnias, e que podem ser associadas a expressões faciais humanas. Ortony [8] explica que cada autor possui a sua classificação sobre quais são as emoções básicas, porém o trabalho melhor aceito é o de Ekman [9] que relizou um estudo aprofundado sobre a relação entre o deslocamento dos músculos faciais e o estado emocional humano.

Ekman descreveu seis emoções como sendo as básicas, são elas: raiva (ou cólera), aversão (ou nojo), medo, felicidade (ou alegria), tristeza e surpresa. Vale ressaltar que, posteriormente, Ekman incluiu o desprezo (ou desdém) à sua lista de emoções básicas porém neste trabalho esta emoção não foi considerada por possuir apenas 18 sequências de imagens contendo exemplos dessa emoção na base de dados Cohn-Kanade [2], escolhida dentre outras (por exemplo a *MMI* [10] ou a *JAFEE* [11]) para execução dos experimentos desse trabalho. Complementando a tese, Ekman e Frriesen construíram o *Facial Action Coding System* (*FACS*) [12], ou Código de Anotação da Ação Facial, com aproximadamente 44 independentes e inter-relacionadas *Action Units - AU*, ou Unidades de Ação, que possibilita classificar os movimentos faciais.

A Tabela 1.1 descreve sucintamente as deformações na face humana durante a manifestação das emoções básicas de Ekman, e a Figura 1.1 mostra exemplos das expressões correspondentes com imagens utilizadas nesse trabalho.

(a) Neutra (b) Raiva (c) Aversão (d) Medo (e) Felicidade (f) Tristeza (g) Surpresa

Figura 1.1: Emoções básicas sugeridas por Ekman [1] demonstradas com imagens da base de dados Cohn-Kanade [2]

Ao longo dos últimos vinte anos, diversos trabalhos tiveram como objetivo reconhecer expressões faciais humanas [13], [14] e [15], no entanto as máquinas ainda não conseguem compreender expressões faciais com a mesma eficiência dos seres humanos. Os sistemas de reconhecimento de expressões faciais podem ser divididos em duas categorias: aqueles que trabalham com imagens estáticas [16], [17] e os que trabalham com sequências de ima-

Tabela 1.1: Expressões Faciais Básicas [1]

Expressão	**Descrição**
Neutra	Todos os músculos da face estão relaxados. As pálpebras estão tangenciando a iris. A boca está fechada e os lábios estão em contato.
Raiva	As sobrancelhas ficam voltadas baixo e juntas. Há tensão na região dos olhos que podem estar mais abertos ou menos abertos que o normal. Os lábios são tensionados e ficam mais finos.
Aversão	As bochechas são empurradas para cima, fazendo com que as pálpebras inferiores se elevem um pouco e estreitem os olhos. Podem aparecer rugas na ponta do nariz.
Medo	As sobrancelhas são levantadas e aproximadas. A pálpebra superior é levantada e a inferior é tensionada. Os cantos dos lábios são puxadas para trás.
Felicidade	As bochechas são levantas e os olhos se fecham parcialmente. A boca se abre mostrando os dentes e os cantos dos lábios são puxadas para trás.
Tristeza	Os cantos internos das sobrancelhas se aproximam. Os cantos internos dos lábios ficam voltados para baixo.
Surpresa	Os lábios relaxam e a boca fica semi-aberta. As sobrancelhas são levantadas e os ficam abertos.

gens [18]. A principal diferença entre eles é a cobertura do vetor de características, pois as abordagens que se baseiam em imagens estáticas utilizam somente a informação de um quadro para classificar uma expressão facial, enquanto que os métodos que se baseiam em uma sequência de imagens utilizam a informação da evolução temporal da deformação dos músculos faciais para reconhecer expressões em um ou mais quadros.

Tian [19] argumenta que a contração dos músculos faciais produz mudanças tanto na aparência quanto na coloração da pele humana, as quais podem ser classificadas em dois tipos: permanentes e transientes. Exemplos de características que fazem parte do primeiro grupo são: localização espacial dos olhos, pálpebras, lábios e sobrancelhas. Mudanças na forma geométrica dos elementos faciais também fazem parte deste grupo. Exemplos do segundo grupo são: a aparição de rugas, linhas verticais e horizontais na testa, sulcos faciais ou qualquer outra característica que não pode ser observada quando a face está completamente relaxada mas aparece durante a manifestação de uma expressão.

Podemos observar que a ocorrência das características que fazem parte do grupo das transientes varia bastante dentre diferentes indivíduos. Por exemplo, uma pessoa idosa pode facilmente apresentar rugas ao redor da face mesmo quanto está com a face relaxada ou rugas entre os olhos podem não ocorrer em alguns indivíduos. Então, considerando que as características transientes podem não se manifestar de maneira similar em pessoas diferentes, utilizar características desse grupo em um sistema de reconhecimento de expressões faciais pode não resultar em um mecanismo genérico e independente de indivíduo. Sendo assim, este trabalho utiliza somente as características do grupo das permanentes, empregando-as como entrada para o cálculo de distâncias como: distância euclidiana entre os olhos ou distância euclidiana entre os lábios e sobrancelhas, aqui também chamadas de características geométricas.

1.3 Objetivos

O objetivo geral dessa dissertação é elaborar um estudo sobre reconhecimento automático de expressões faciais humanas baseando-se na geometria facial. Os objetivos específicos são:

- Investigar diferentes métodos de extração e seleção de características para o reconhecimento de expressões faciais humanas.

- Propor um método de reconhecimento de expressões faciais que seja totalmente automatizado.

- Implementar um protótipo de programa de computador que possibilite avaliar experimentalmente o método proposto;

1.4 Organização do Documento

O restante desse documento está organizado em mais quatro capítulos. O capítulo 2 apresenta o resultado da revisão bibliográfica mostrando diferentes abordagens de extração de características faciais e trabalhos relacionados. Em seguida, o capítulo 3 apresenta o método proposto bem como duas abordagens de seleção de características para o reconhecimento de expressões faciais. O capítulo 4 descreve os resultados de desempenho das abordagens propostas, comparando com outros trabalhos relacionados. Finalmente, o capítulo 5 apresenta conclusões e trabalhos futuros.

Capítulo 2

Revisão Bibliográfica

Durante o levantamento bibliográfico foram encontrados diversos trabalhos que tratam do reconhecimento de expressões faciais humanas. Revisões bibliográficas completas podem ser encontradas em: [13], [14], [15], [20], [21], [22] e [23].

Segundo Liu *et al.* [16], os sistemas de reconhecimento de expressões faciais basicamente utilizam o procedimento de treinamento em três etapas: extração de características, seleção de características e construção do classificador. A primeira é responsável por obter todas as características relacionadas a variação da expressão facial. Apesar de muitas vezes a etapa seguinte estar embutida na primeira, a esta é responsável por escolher as melhores características para representação da expressão facial. Nesta fase o objetivo é minimizar a variação intra-classe e maximizar a variação inter-classe de expressões [17]. No entanto, no contexto de expressões faciais, minimizar a variação intra-classe é um desafio porque imagens que contêm pessoas diferentes manifestando a mesma expressão facial podem ser bem diferentes no espaço dos pixels. Por sua vez, maximizar a variação inter-classe também é uma tarefa desafiadora pois imagens de uma mesma pessoa manifestando diferentes expressões faciais podem ser bem parecidas no espaço dos pixels [24]. Por último, um classificador (ou uma combinação de classificadores como em [25]) é utilizado para inferir a expressão facial dadas as características selecionadas.

De forma geral, as abordagens de reconhecimento de expressões faciais variam na forma como é realizada a extração e seleção de características e no método de classificação utilizado. Fasel e Luettin [13] argumentam que a obtenção e o rastreamento das características que representam a deformação facial são etapas cruciais para a análise das expressões faci-

ais. Levando em consideração que não é objetivo avaliar diferentes métodos de classificação, este trabalho enfoca nos mecanismos de extração e seleção de características faciais, somente avaliando a performance dos diferentes conjuntos de características utilizados sem alterar o algoritmo de classificação de padrões, no caso, o escolhido foi a Máquina de Vetores de Suporte *Support Vector Machines - SVM* por lidar bem com problemas pouca quantidade de exemplos de treinamento.

2.1 Métodos de Extração e Seleção de Características Faciais

Durante a pesquisa bibliográfica, vários sistemas de reconhecimento da expressão facial foram encontrados seguindo diferentes abordagens de extração e seleção das características faciais [21]. No entanto, de uma forma geral, eles estão agrupados em duas categorias: métodos de extração baseados na aparência da face e métodos baseados em um modelo da face humana.

Os métodos baseados na aparência da face buscam processar imagens contendo a face humana, ou regiões da face humana, a fim de identificar mudanças locais nos níveis de cinza durante a manifestação de uma expressão. As técnicas que pertencem a esse grupo se concentram em observar alterações na textura facial, em alguns casos segmentando a face em regiões, e utilizam descritores de textura como *Wavelets* de Gabor [26] ou Padrões Locais Binários - LBP [27] para extrair características das expressões faciais. Os algoritmos baseados nesta abordagem possuem a vantagem de conter efetivamente a informação da expressão manifestada mas normalmente precisam ser acompanhados de algum método de redução de dimensionalidade do vetor de características como Análise de Componentes Principais - PCA (*Principal Component Analysis*) ou Análise de Discriminantes Lineares - LDA (*Linear Discriminant Analysis*) a fim de reduzir a quantidade de cálculos necessários para a posterior classificação. Lillewort *et al.* [28], por exemplo, convoluiu a face detectada em uma imagem em um banco de 72 filtros de Gabor com oito orientações e nove frequências, onde a saída de cada filtro é considerada um atributo. Em seguida, todas as 72 saídas são apresentadas a uma máquina de aprendizado SVM que classifica as *Action Units* - AU [12]. Por fim, a saída da SVM contendo as AUs encontradas são levadas para um classificador de Regressão

Logística para o reconhecimento da expressão facial.

Os métodos que se baseiam em modelos da face humana normalmente consistem em utilizar modelos para representar as estruturas faciais, e então determinar o deslocamento e a deformação dos seus componentes. Algumas técnicas que fazem parte desse grupo, como as apresentadas no Capítulo 3, são chamadas de abordagens geométricas pois, além de utilizar um modelo do formato da face, utilizam as mudanças nas distâncias entre pontos de controle causadas pela variação da expressão para determinar as deformações dos músculos da face. As abordagens de reconhecimento que se baseiam em características geométricas dispõem da vantagem de possuírem um processo relativamente simples e fácil de reconhecimento, de necessitarem de pouco espaço de memória já que expressam cada imagem em um conjunto de vetores de características culminando em um número menor de cálculos para a etapa de classificação, e sofrerem pouco impacto sobre diferenças de iluminação. No entanto, esta abordagem também possui a desvantagem de não possuir a informação da face completa se traduzindo em dificuldades na detecção de mudanças sutis.

Saeed *et al.* [29] realizou a inferência da expressão facial utilizando características baseadas na localização de oito pontos de controle, calculados através de uma técnica de Modelos Pontuais de Distribuição - PDM (*Point Distribution Models*), representando a forma e a localização de três componentes faciais: olhos, sobrancelhas e boca. Em seguida, os autores derivaram seis características geométricas e realizaram um experimento levando em consideração a expressão neutra e outro desconsiderando-a. Quando não foi considerada a expressão neutra, foi encontrado o resultado de 83% de acurácia. Ao adicionar a expressão neutra, a taxa de reconhecimento caiu para $73,6\%$ utilizando SVM como abordagem de classificação de padrões. Tang *et al.* [30] utilizou um Modelo de Aparência Ativa - AAM para extrair 63 pontos de controle derivando deles quatro características efetivas para o reconhecimento de expressões. Eles obtiveram 88% de taxa de reconhecimento ao classificar 4 das emoções básicas.

2.2 Trabalhos Relacionados

Sabendo da diversidade de técnicas que podem ser utilizadas ao longo das etapas de um sistema reconhecedor de expressões faciais, esta seção lista alguns trabalhos relevantes que

utilizaram a base de dados *Cohn-kanade*, tornando possível a comparação direta dos resultados obtidos nesse trabalho, já que todos os experimentos foram realizados com ela.

Shan *et al.* [17] investigou o impacto da resolução da imagem na taxa de reconhecimento dos sistemas de reconhecimento de expressões faciais e concluiu que as abordagens baseadas em características geométricas não trabalham bem com baixas resoluções enquanto que as abordagens que se baseiam na aparência do objeto, como *wavelets* de Gabor e padrões locais binários (*Local Binary Patterns - LBP*), não sofrem tanto. Além disso, os autores também executaram um estudo profundo que, no melhor cenário, alcançou $95,1\%$ utilizando o classificador SVM e LBP como extrator de características. Este trabalho representa o atual estado da arte na classificação de expressões faciais utilizando a base de dados Cohn-kanade [2].

Devido ao pequeno número de classes e vetor de características de alta dimensionalidade, SVM é bastante adequando para os sistemas de reconhecimento de expressões faciais. Hsieh *et al.* [31] utilizou Adaboost e ASM (*Active Shape Model*) [6], para identificar a face humana e localizar os componentes faciais. Em seguida, os autores utilizaram filtros de Gabor e o algoritmo de detecção de bordas LoG (Laplace of Gaussian) para propor "características faciais semânticas". Por fim, SVM foi utilizado para classificar seis expressões faciais (a expressão Tristeza foi desconsiderada) chegando a uma média de $94,5\%$ de taxa de reconhecimento na base de dados Cohn-kanade [2].

Michel *et al.* [32] definiram 22 pontos de controle para o reconhecimento de expressões faciais. O movimento de todos os pontos de controle desde a expressão neutra até o auge da manifestação da expressão foi medido como um vetor de características. Chen *et al.* [33] utilizou o deslocamento dos pontos de controle e diferenças de textura entre a expressão neutra normalizada e imagens com expressões faciais para o reconhecimento. O vetor de atributos obtido continha características geométricas, com 42 dimensões, e características de textura com 21 dimensões. A taxa de reconhecimento média encontrada foi de 95% (desconsiderando a expressão Aversão) utilizando o classificador SVM e a base de dados Cohn-kanade [2].

O sistema proposto por Kotsia *et al* [34], encontrou $99,7\%$ de taxa de reconhecimento utilizando sequências de imagens. No entanto, o método proposto possui a desvantagem de que inicialmente é necessário que o usuário indique manualmente a localização dos de controle em uma face manifestando a expressão neutra e, em seguida, o sistema analisa a

sequência de imagens até que a expressão facial seja manifestada por completo, desde a expressão neutra até o ápice da manifestação da expressão.

Capítulo 3

Sistema de Reconhecimento de Expressões Faciais

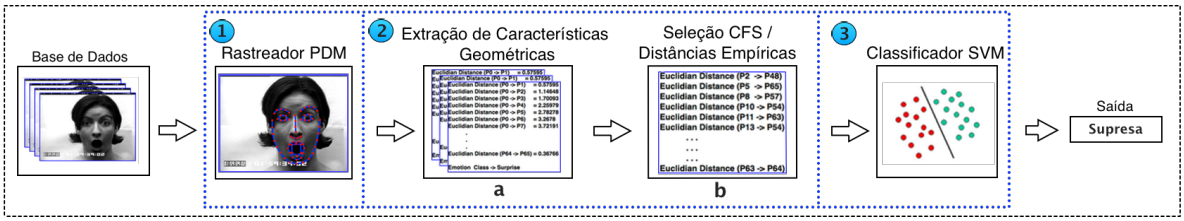


Figura 3.1: Visão geral do método proposto.

Neste capítulo, um método baseado na geometria facial para o reconhecimento de expressões faciais é apresentado. Uma visão geral do método proposto está ilustrada na Figura 3.1.

O sistema proposto é composto de três fases principais: rastreador PDM, extração e seleção de características e classificação da expressão. Na primeira fase, o sistema rastreia os pontos de controle definidos neste trabalho. Na fase seguinte, o sistema extrai a distância Euclidiana de todos os pontos de controle e seleciona as características mais relevantes. Nesta etapa são apresentadas duas propostas de seleção de características. Na última fase, o sistema classifica a expressão em uma das emoções básicas, apresentadas na Tabela 1.1, utilizando uma máquina de vetores de suporte com *kernel RBF - Radial Basis Function*. O restante desse capítulo detalha cada uma das fases.

3.1 Modelos Pontuais de Distribuição

Por ser um objeto não rígido, isto é, que pode sofrer mudanças no formato e na aparência dos seus componentes, a face humana requer um método robusto para rastreá-la com eficiência. Além disso, para o reconhecimento de expressões em imagens digitais, é preciso lidar também com diferenças de translação, rotação e escala ou de problemas de oclusão parcial de algum de seus elementos devido a presença de algum componente não estrutural como: barba, bigode, chapéu, óculos e etc. Para resolver esse problema, os Modelos Pontuais de Distribuição, do inglês *Point Distribution Models - PDM*, buscam ajustar um modelo deformável do formato do objeto a uma nova instância. Nesta abordagem, o formato do objeto é descrito através de um conjunto de n pontos, chamados de pontos de controle (ou *landmarks*, ou pontos fiduciais), podendo representar tanto as características internas quanto as externas de um objeto, isto é, pode-se representar o contorno da face e os seus elementos internos como olhos, boca, nariz, lábios, sobrancelhas etc.

Neste trabalho foi utilizado um conjunto de 66 pontos de controle, que foram obtidos através da junção dos pontos em comum entre os 76 pontos de controle sugeridos pela base de dados MUCT[4] e os disponíveis na base de dados *Cohn-Kande* [2], que neste trabalho se mostraram suficientes para descrever o movimento de todas as características permanentes da face. A disposição dos pontos de controle utilizados pode ser verificada na Figura 3.2.

O primeiro passo para construir um PDM é possuir uma base de dados, geralmente manualmente marcada, contendo exemplos dos diferentes formatos que o objeto pode apresentar. Em seguida, deve-se realizar uma análise estatística que possibilite descrever os diferentes formatos que o objeto possa assumir. Seguindo esse pensamento, Cootes *et al.* [6] propôs os Modelos de Forma Ativa (ou em inglês, *Active Shape Models - ASM*), em que primeiramente alinham-se todos os exemplos da base de dados aplicando uma Análise Generalizada de Procrustes [35], removendo variações de rotação, transalação e escala, e então constrói-se um modelo linear da forma do objeto aplicando uma Análise de Componentes Principais - PCA. Isto significa uma forma s pode ser expressada através de uma forma base s_0 acrescida de uma combinação linear de m variações de sua forma s_i, ilustrado na equação (3.1), onde os coeficientes $p = (p_1, ..., p_m)^T$ são os parâmetros da forma. A Figura 3.3 ilustra a ideia.

Em etapa posterior, para ajustar o modelo do formato do objeto a um nova instância

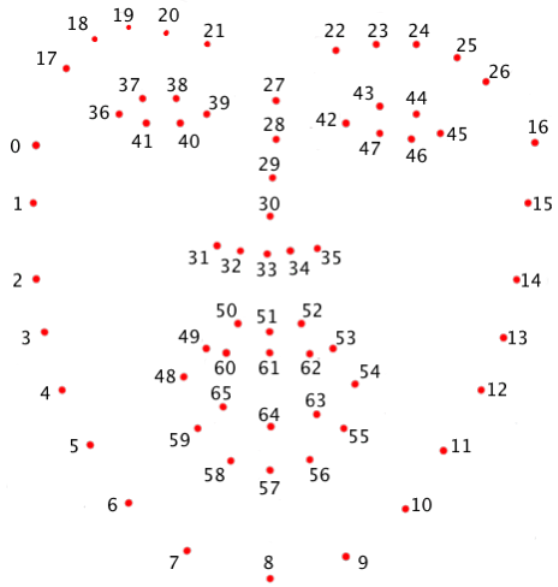


Figura 3.2: Conjunto de pontos utilizados para representar os componentes faciais.

do objeto, inicialmente faz-se uma busca na área em torno de cada ponto de modelo (na primeira estimativa considera-se a forma média do objeto) a fim de estimar o deslocamento de cada ponto. Por fim, o modelo da forma do objeto é então concatenado as localizações estimadas a fim de aplicar restrições globais de formato do objeto para impossibilitar que formas inválidas do objeto sejam geradas.

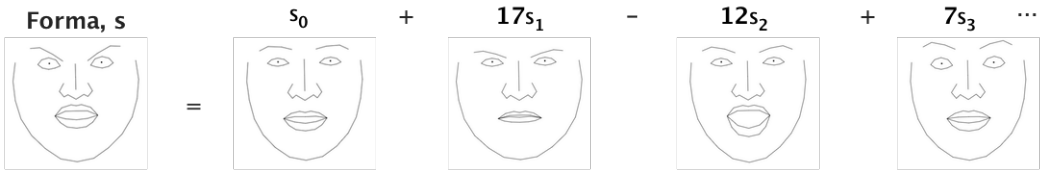


Figura 3.3: Uma forma s pode expressada através de uma forma base s_0 acrescida de uma combinação linear das variações da sua foram s_i.

$$s = s_0 + \sum_{i=1}^{m} p_i s_i \tag{3.1}$$

Após a proposta inicial de Cootes *et al.*[6], as abordagens que ajustam os pontos de controle baseando-se em um modelo deformável aplicando restrições aos formatos que podem ser estimados através de uma análise estatística, é referenciado na literatura como Modelo

Pontual de Distribuição.

Para uma melhor compreensão do leitor, o apêndice A descreve de maneira detalhada os passos para a construção de um Modelo Pontual de Distribuição como proposto originalmente. As seções 3.1.1 e 3.1.2 discorrem sobre dois tipos de PDM utilizados nesse trabalho: o Modelo de Aparência Ativa (*Active Appearance Model* - AAM) e o Modelo de Restrições Locais (Constraint Local Model - CLM).

3.1.1 Modelos de Aparência Ativa

Após o ASM, Cootes *et al.* propôs o *Active Appearance Model* - AAM ou Modelo de Aparência Ativa [36]. Diferentemente dos ASMs que buscam ajustar a posição de cada ponto do modelo deformável diretamente na imagem de entrada, o AAM ajusta os parâmetros dos modelos de modo a maximizar a correspondência entra uma instância do modelo e uma imagem de entrada.

Por ser um PDM, os AAM são construídos a partir de um conjunto de imagens manualmente marcadas, e extraindo do exemplos dois modelos: o Modelo de Forma e o Modelo de Aparência ou Textura.

Modelo de Formas do AAM

Matematicamente, o modelo de formas do AAM segue a mesma definição do modelo de formas do ASM, em que a forma s é composta por uma forma base s_0 acrescida de uma combinação linear de n vetores de formas s_i, onde os coeficientes p_i são os parâmetros do modelo da forma (ver equação (3.1)). A Figura 3.4 apresenta um exemplo de um modelo de formas de um AAM em que a forma base se apresenta à esquerda, e à direita os primeiros três vetores de forma s_1, s_2 e s_3. Observa-se que, diferentemente do ASM, o modelo de forma do AAM refere-se a uma malha triangulada através da triangulação de Delaunay [37] utilizando os pontos de controle como referência. Finalmente, os parâmetros do modelo de forma do AAM, dados pelo vetor $p = (p_1, p_2, ..., p_n)^T$, são utilizados como atributos de entrada para um algoritmo de reconhecimento de padrões.

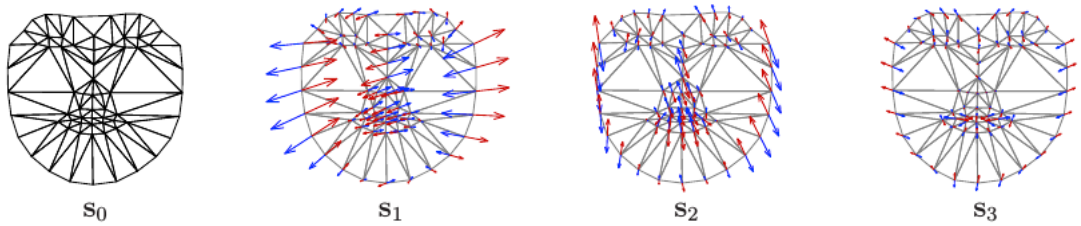


Figura 3.4: Modelo linear da forma de um AAM.

Modelo de Aparência

A aparência de um AAM está definida dentro da malha base s_0. Admitindo que s_0 também denote o conjunto de pixels $x = (x, y)^T$ dentro da malha base s_0, a aparência de um AAM é uma imagem $A(x)$ definida sobre o pixels $x \epsilon s_0$. Similar ao modelo de forma, a imagem de aparência $A(x)$ pode ser expressa através de uma aparência base $A_0(x)$ mais uma combinação linear de m imagens de aparência $A_i(x)$:

$$A(x) = A_0(x) + \sum_{i=1}^{m} \lambda_i A_i(x) \tag{3.2}$$

onde os coeficientes λ_i são os parâmetros do modelo de aparências. A aparência base $A_0(x)$ corresponde à imagem média, e as imagens $A_i(x)$ às m autoimagens associadas aos m maiores autovalores.

Uma vez obtidos os modelos de forma e de aparências, realiza-se a etapa de instanciação de um modelo AAM.

Instanciação do Modelo AAM

Supondo inicialmente conhecidos os parâmetros $p = (p_1, p_2, ..., p_n)^T$, do modelo de forma do AAM, a Equação (3.1) pode ser utilizada para gerar a forma s. De maneira equivalente, se conhecidos os parâmetros do modelo de aparência $\lambda = (\lambda_1, \lambda_2, ..., \lambda_n)^T$, a aparência $A(x)$ definida no interior da malha base s_0 pode ser gerada. Deste modo, a instância do modelo do AAM com parâmetros de forma p e parâmetros de aparência λ é criada levando a aparência A desde a malha base s_0 até o modelo de forma s. Este procedimento é ilustrado na Figura 3.5 para um conjunto de valores p e λ.

Em particular, o par de formas s_0 e s definem uma deformação afim por trechos desde s_0 até s. Esta deformação afim consiste no seguinte procedimento:

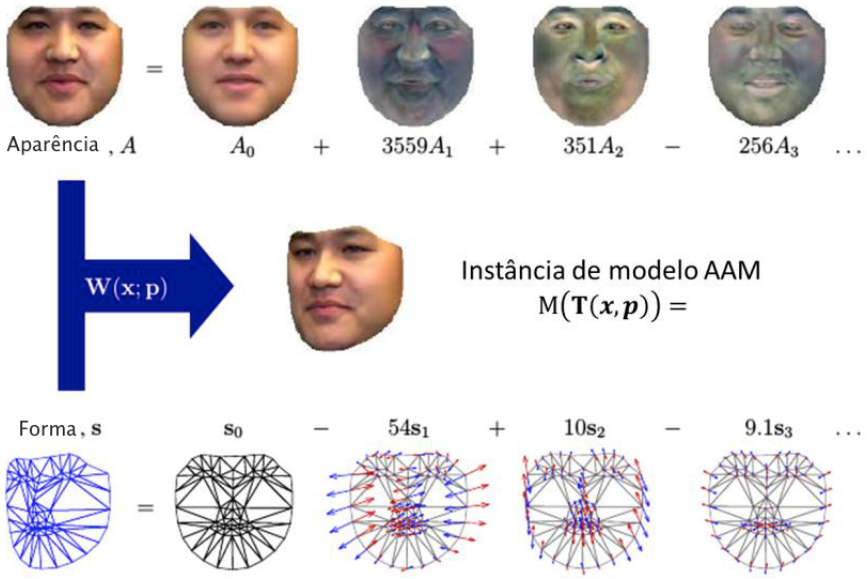


Figura 3.5: Exemplo de instanciação de um AAM. Na parte superior da imagem tem-se o modelo de aparência. Na parte inferior, o modelo da forma. No centro, a concatenação dos modelos gerando uma instância denotada por $M(T(x,p))$ que correspondente a imagem real $I(x)$.

Cada vértice de cada triângulo em s_0 está associado a um vértice de um triângulo particular em s, por meio de uma transformação geométrica (translação, rotação e escala). Esta correspondência determina uma deformação afim única que vai de um triangulo em s_0 para outro em s, ou seja, os vértices do primeiro triângulo mapeiam em vértices no segundo. Deste modo, identificada a correspondência entre cada um dos triângulos, se procede com o conjunto de pixels que compõem as duas imagens, isto é, as aparências das duas imagens são alinhadas. Assim, para cada pixel identificado por seu vetor de coordenadas $x = (x, y)$ em s_0 deve-se primeiro determinar o par de triângulos associados em s_0 e s, e aplicar-lhe a função de deformação afim correspondente entre esses dois triângulos. Finalmente, o alinhamento ocorre mediante a aplicação de deformações afim específicas aos pixels que compõem cada triângulo, denotadas por $T(x,p)$. Desta maneira, a instância final do modelo AAM é calculada deformando a aparência A de s_0 para s, utilizando a relação $T(x,p)$. Este processo é representado pela seguinte equação:

$$M(T(x,p)) = A(x) \tag{3.3}$$

onde M é uma imagem $2D$ de tamanho e forma apropriados para comportar a instância do modelo. Podemos interpretar a equação 3.3 da seguinte maneira: Dado um pixel x em s_0, seu correspondente em s é dado por $T(x, p)$.

Ajuste de um Modelo AAM a uma Imagem de Entrada

A partir de uma imagem de entrada $I(x)$, contento o objeto modelo por um AAM, é preciso determinar os parâmetros de forma p e de aparência λ datas equações 3.1 e 3.2 respectivamente. Valores ótimos destes parâmetros culminam no melhor ajuste da imagem de entrada $I(x)$ à correspondente instância $M(T(x, p)) = A(x)$. Portanto, para ajustar um modelo AAM é preciso determinar os valores de p e λ que minimizam a discrepância entre a imagem de entrada $I(x)$ e $M(T(x, p))$.

Considerando que x é o vetor de coordenadas de um pixel em s_0, o pixel correspondente na imagem de entrada I está na coordenada $T(x, p)$. Considerando que no pixel x o AAM tem aparência $A(x) = A_0(x) + \sum_{i=1}^{m} \lambda_i A_i(x)$ e no pixel $T(x, p)$, a imagem de entrada tem intensidade $I(T(x, p))$. Sendo assim, para minimizar a discrepância entre $A(x)$ e $I(T(x, p))$ para o conjunto de pixels em s_0, deve-se minimizar a equação 3.4, com respeito aos parâmetros de forma p e de aparência λ, onde a soma é realizada sobre todos os pixels x na malha s_0.

$$\sum x \epsilon s_0 \left[A_0(x) + \sum_{i=1}^{m} \lambda_i A_i(x) - I(T(x, p)) \right]^2 \tag{3.4}$$

Diversas técnicas foram propostas para resolver, de maneira eficiente, o problema da minimização da equação (3.4). Matthews e Baker [38] resumem as principais técnicas desenvolvidas para esse propósito. No AAM utilizado nessa dissertação foi utilizado a técnica conhecida como *Inverse Compositional Image Alignment* sugerida por Matthews e Baker [38].

3.1.2 Modelos de Restrições Locais

Os Modelos de Restrições Locais ou *Constrained Local Models* (CLM) se caracterizam como uma evolução do AAM pois ao invés de ajustar a textura de todo o objeto, busca-se ajustar a textura ao redor de cada um dos pontos de controle que compõe o modelo da

forma do objeto. Especificamente, os CLM precisam de duas fontes de informação para operar: um modelo individual da aparência de cada um dos pontos de controle, chamado de modelo local de fragmentos (ou *patches*, em inglês) e um modelo do formas do objeto. A Figura 3.6 ilustra os modelos construídos nessa abordagem.

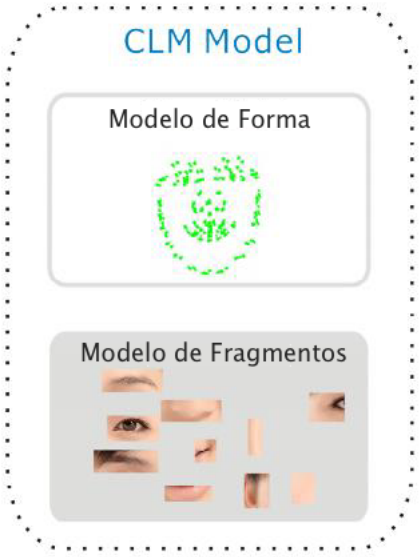


Figura 3.6: Representação do modelo de formas e o modelo de fragmentos de um CLM.

Assim como no ASM e no AAM, os modelos são gerados a partir de um conjunto de imagens de faces etiquetadas de forma manual. A Figura 3.7 mostra uma visão geral das etapas de um CLM.

Assim como os ASMs, a abordagem os Modelos de Restrições Locais constroem um modelo de formas clássico de um PDM, conforme descrito na equação (3.1).

Modelo de Fragmentos

Para construir um modelo de fragmentos, deve-se utilizar descritores de textura especializados em cada ponto do modelo. Para isso, Cristinacce *et al.* [39] sugerem treinar um conjunto de n classificadores do tipo SVM [40], ou seja, um classificador para cada fragmento, especializado em identificar as características de textura na vizinhança de cada ponto de controle. Para treinar a SVM de cada um dos pontos, o autor seleciona M exemplos positivos e M exemplos negativos (exemplos que fazem referência a outros pontos de controle escolhidos de forma aleatória). Vale ressaltar que um exemplo é definido como uma janela de pixels

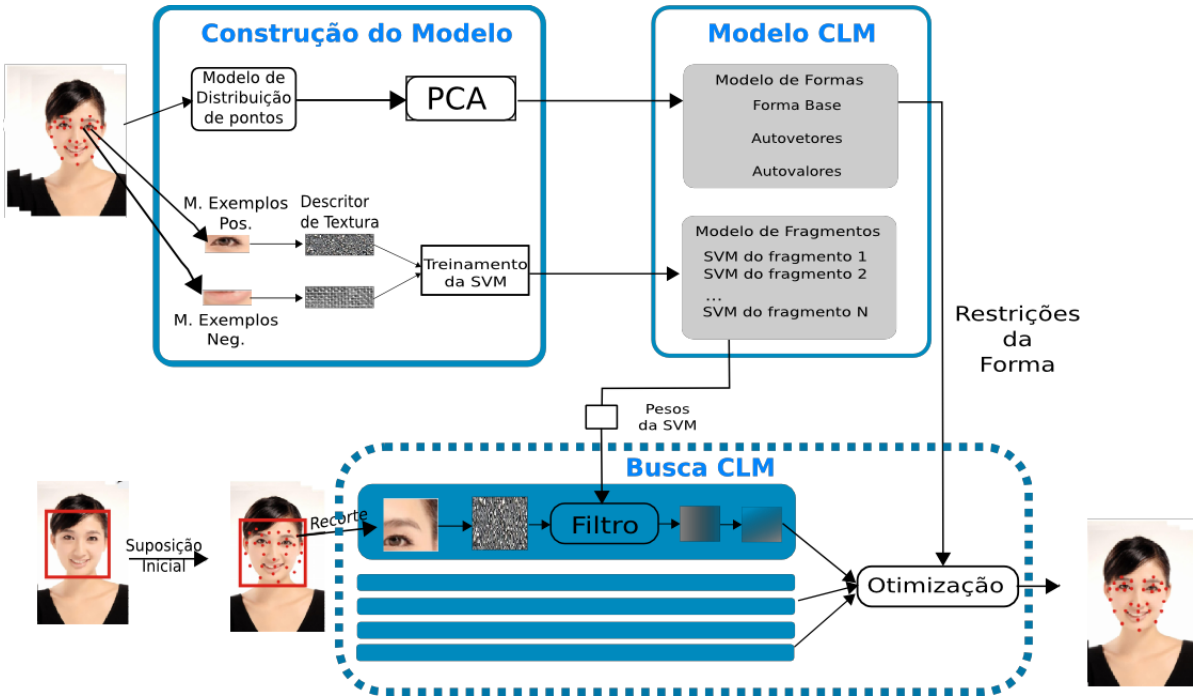


Figura 3.7: Visão geral do funcionamento de um CLM.

centrada no ponto de controle de interesse, cujas dimensões variam conforme a resolução das imagens de treinamento. A seção 3.3.1 descreve sucintamente o funcionamento de uma SVM.

Ajuste do Modelo CLM

Inicialmente, antes detectar os pontos de controle, deve-se estimar a localização do objeto na imagem. Para este propósito, pode-se utilizar um detector de objetos genérico como Viola-Jones [41]. Em seguida, deve-se realizar uma suposição inicial quanto à localização dos pontos de controle na imagem de entrada. Geralmente, para esse propósito, é utilizada a forma base s_0. Então, para cada ponto de controle é selecionada uma janela de pixels de dimensão maior do que a utilizada para montar o modelo de fragmentos . A matriz com indicadores da localização dos pontos de controle fornece os indícios da localização estimada do ponto de controle procurado. A esta altura, a informação fornecida pelo modelo de fragmentos é comparada com todas as possíveis variações do modelo de forma para que, caso necessário, o ponto de controle seja ajustado para uma coordenada que possivelmente seja mais próxima da posição desejada. Este processo é repetido iterativamente até o algoritmo

convergir ou até um determinado número de iterações. A implementação do CLM utilizada neste trabalho foi a sugerida por Saragih [42].

3.2 Extração e Seleção de Características

De posse do formato da face descrito através de um conjunto de pontos que representam a geometria facial, deve-se então selecionar as características que maximizam o limite de decisão entre as classes quando as distâncias entre os pontos de controle mudam, isto é, o objetivo é identificar os atributos de carregam informação para o classificador e eliminar atributos que não fornecem tanta informação ou são redundantes.

Segundo Katti *et al.* [43], as técnicas de seleção de características podem ser classificadas de diferentes maneiras, uma delas está relacionada a sua relação com o algoritmo de indução[1]. Para avaliar a qualidade e a performance de um subconjunto de atributos, três abordagens são comumente utilizadas: embutida, baseada em filtro e baseada em *wrapper*.

No primeiro caso, a seleção do subconjunto é embutida ou integrada ao algoritmo de indução, como por exemplo, as árvores de decisão. Já a abordagem baseada em filtro, trabalha independente de qualquer algoritmo de indução em que, em uma etapa de pré-processamento, é utilizado um filtro sobre o conjunto original de atributos. As técnicas que pertencem a esse grupo, por exemplo, verificam a correlação entre os atributos e normalmente são mais rápidos que as técnicas do grupo anterior. Por último, as abordagens baseadas em *wrapper* argumentam que o *viés* de um algoritmo de indução deve ser levado em consideração na etapa de seleção de características. Sendo assim, para cada subconjunto de atributos possível, o algoritmo de indução é consultado e o subconjunto que tiver maior redução da taxa de erro é normalmente selecionado.

Neste trabalho, para encontrar o melhor subconjunto de características, primeiramente todas as distancias Euclidianas D entre os pontos de controle que descrevem o formato da face humana (ilustrado na Figura (3.2)) foram calculados através da equação 3.5. Sabendo

[1]Em Aprendizagem de Máquina - AM, computadores são programados para aprender com a experiência passada. Para tal, empregam um princípio de inferência denominado indução, no qual se obtêm conclusões genéricas a partir de um conjunto particular de exemplos. Assim, algoritmos de AM aprendem a induzir uma função ou hipótese capaz de resolver um problema a partir de dados que apresentam instâncias do problema a ser resolvido. [43]

que são considerados 66 pontos de controle, tem-se $\binom{66}{2} = 2145$ distâncias a serem consideradas.

A fim de reduzir a dependência de diferentes tamanhos de face, problemas de escala e variações de translação, propõe-se normalizar todos os pontos P_i utilizando a equação (3.6) onde Dn é o coeficiente de normalização calculado utilizando a distância entre o canto esquerdo do olho direito e o canto direito do olho esquerdo, conforme a equação (3.7). Esta distância foi selecionada porque durante os experimentos observou-se que ela sofria pouca alteração durante a deformação dos músculos da face.

$$D_{real}(P_\imath, P_\jmath) = \|P_{\imath_{normalizado}} - P_{\jmath_{normalizado}}\| \tag{3.5}$$

$$P_{\imath_{normalizado}} = P_\imath / Dn \tag{3.6}$$

$$Dn = D(P_{42}, P_{39}) \tag{3.7}$$

Neste trabalho, duas abordagens de seleção de características são propostas. A primeira foi baseada no ponto de vista humano cujas características foram obtidas através de um experimento empírico. A segunda abordagem aplica o algoritmo de seleção de atributos por correlação (*Correlation Features Selection - CFS*), ambas são abordagens baseadas em filtro.

3.2.1 Distâncias Empíricas

Para selecionar as distâncias a serem utilizadas na etapa de classificação de características, inicialmente, uma pesquisa experimental foi realizada apresentando para um grupo de dez pessoas 100 imagens contendo somente os pontos de controle correspondentes a faces extraídas da base de dados *Cohn-Kanade* manifestando uma emoção básica. Exemplos de imagens utilizadas no experimento podem ser observadas na Figura 3.8. Empiricamente, quando os participantes acertavam a expressão mostrada, era questionado quais foram os fatores que foram levados em consideração para que eles chegassem as suas conclusões. O propósito deste estudo foi identificar, de um ponto de vista humano, quais pontos de controle ou distâncias entre eles podem ser levadas em consideração para classificar expressões faciais.

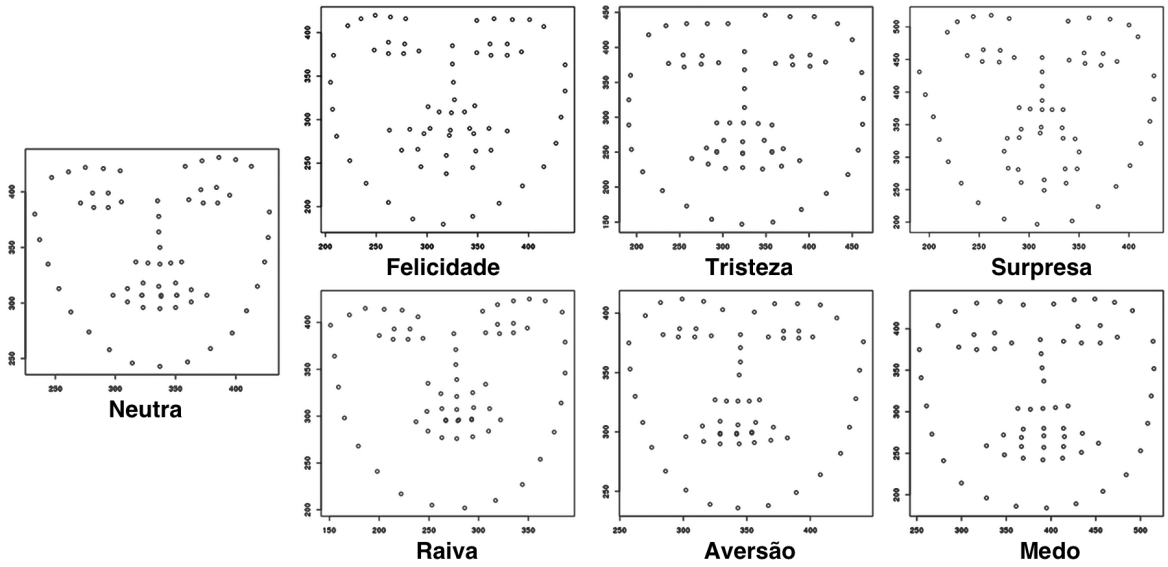


Figura 3.8: Exemplos de imagens utilizadas em uma pesquisa experimental realizada com o objetivo de identificar quais as distâncias mais relevantes segundo a visão humana para a classificação de expressões faciais.

Ao realizar o experimento mostrando aos voluntários somente a disposição dos pontos de controle, observou-se que as expressões surpresa e felicidade foram facilmente reconhecidas pelos voluntários através da distância entre as pálpebras, entre os lábios e entre os cantos da boca. As expressões neutra, aversão e raiva, apesar de causarem confusão entre si, muitas vezes eram distinguidas através da distância entre as sobrancelhas e a altura das sobrancelhas e do queixo. A tristeza e o medo são as que mais geravam dúvidas entre os participantes, quando identificadas, foi levada em consideração a relação entre o canto do olho e o canto da boca. Em média, a taxa de acertos dos voluntários foi de 53%.

Baseado no experimento descrito acima, são propostas sete características para classificar expressões faciais somente utilizando a geometria facial. A descrição das distâncias utilizadas pode ser observada na Tabela 3.1. Vale ressaltar que, assim como sugerido por Soyel [44], para minimizar problemas de medição da distância entre dois pontos, quando possível, é considerado o valor médio dentre distâncias vizinhas, como por exemplo na característica que mede a abertura dos olhos em que temos dois pontos de controle para representar a pálpebra superior (P_{37} e P_{38}) e outros dois para a pálpebra inferior (P_{40} e P_{41}). A Figura 3.9 ilustra as distâncias consideradas.

Tabela 3.1: Distâncias Empíricas utilizadas.

Nome da Distância	**Descrição**
Abertura dos Olhos	$\frac{Dreal(P_{37},P_{41})+Dreal(P_{38},P_{40})}{2}$
Altura da Sobrancelha	$\frac{Dreal(P_{20},P_{38})+Dreal(P_{18},P_{37})}{2}$
Distância entre as Sobrancelhas	$Dreal(P_{21}, P_{22})$
Altura da Boca	$\frac{Dreal(P_{60},P_{65})+Dreal(P_{61},P_{64})+Dreal(P_{62},P_{63})}{3}$
Comprimento da Boca	$Dreal(P_{54}, P_{48})$
Altura do Queixo	$Dreal(P_{30}, P_8)$
Alongamento dos Lábios	$Dreal(P_{36}, P_{48})$

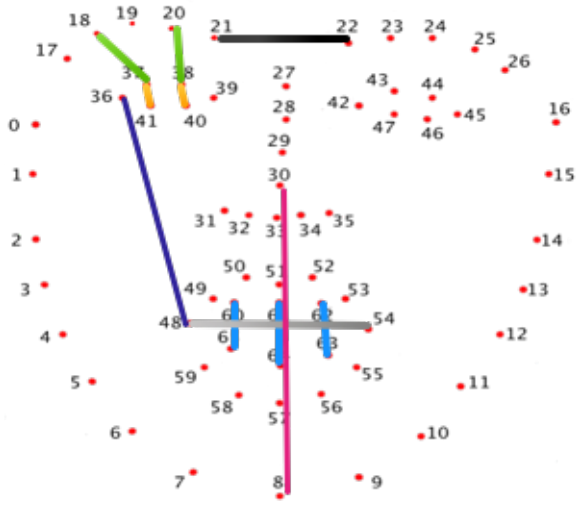


Figura 3.9: Distâncias empíricas consideradas.

3.2.2 Correlation Features Selection - CFS

Assim como a maioria dos métodos de seleção de características, a seleção de características por correlação (*Correlation Features Selection* - CFS) utiliza um algoritmo de busca para explorar o espaço de subconjuntos possíveis em conjunto com uma função que avalia o mérito de cada subconjunto de características. A ideia básica do CFS é que: *Bons subconjuntos possuem características altamente correlacionadas com as classes e não-correlacionados entre si*. Assim, tanto características irrelevantes quanto redundantes podem ser desprezadas. A equação (3.8) formaliza a heurística de avaliação do Mérito G_s de cada subconjunto de características:

$$G_s = \frac{k\bar{r}_{ci}}{\sqrt{k + k(k-1)\bar{r}_{ii}}} \tag{3.8}$$

onde k é a cardinalidade, $\bar{r}_{ci}$ é correlação média entre as características e as classes e $\bar{r}_{ii}$ é correlação média das características entre si. Pode-se notar que o numerador da equação (3.8) indica quanto o conjunto de características é correlacionado com as classes. Por sua vez, o denominador indica quanta redundância há no conjunto. Assim, subconjuntos com maiores valores de mérito são preferíveis.

Para calcular $\bar{r}_{ci}$ e $\bar{r}_{ii}$, o CFS utiliza a incerteza simétrica (IS), definida como:

$$IS(\mathbf{x}_1, \mathbf{x}_2) = 2.0 \left[\frac{GI(\mathbf{x}_1, \mathbf{x}_2)}{H(\mathbf{x}_1) + H(\mathbf{x}_2)} \right] \tag{3.9}$$

em que $(\mathbf{x}_1, \mathbf{x}_2)$ são vetores aleatórios que representam as distâncias entre os pontos de controle pertencentes ao espaço de atributos, isto é, $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$. As funções $GI(\mathbf{x}_1, \mathbf{x}_2)$ e $H(\mathbf{x})$ são, respectivamente, segundo [45], o ganho de informação e a entropia, definas pelas Equações (3.11) e (3.10). A IS assume valores de 0 a 1, indicando o nível de associação entre as variáveis, em que maiores valores indicam maiores níveis de associação.

O cálculo do CFS empregado neste trabalho foi baseado na implementação existente no *software* Weka [46]. Nesta implementação, realiza-se uma discretização de cada um dos atributos, como estabelecido por Fayyad e Irani [47], a fim de estimar, através de frequências relativas, as distribuições de probabilidades usadas no cálculo da medida de entropia. Considerando que um dado atributo x_j tenha sido discretizado em c intervalos, estima-se sua entropia, $H(x_j)$, como em (3.10).

$$H(x_j) = \sum_{i=1}^{c} -p(x_{ji}) log_2 p(x_{ji}) \tag{3.10}$$

$$\begin{aligned} GI &= H(\mathbf{x}_1) - H(\mathbf{x}_1|\mathbf{x}_2) \\ &= H(\mathbf{x}_2) - H(\mathbf{x}_2|\mathbf{x}_1) \\ &= H(\mathbf{x}_1) + H(\mathbf{x}_2) - H(\mathbf{x}_1, \mathbf{x}_2) \end{aligned} \tag{3.11}$$

Após a extração das 2145 distâncias Euclidianas possíveis, foi utilizado o algorítimo CFS com uma busca heurística do tipo O-Melhor-Primeiro (*Best First Search* - BFS) para selecionar as distâncias mais adequadas chegando a um subconjunto com 44 distâncias. A Figura 3.10 ilustra as 44 distâncias obtidas ao avaliar a base de dados Cohn-kanade[2].

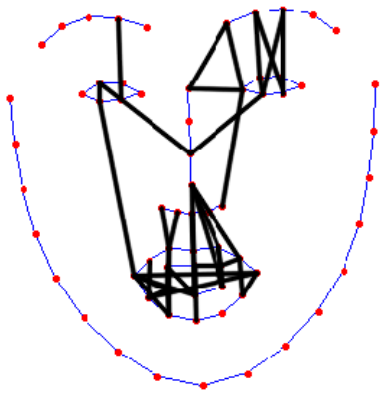


Figura 3.10: Distâncias selecionadas pela seleção de características CFS.

3.3 Classificação de Padrões

Nesse trabalho, foi escolhido como classificador a Máquina de Vetores de Suporte (*Support Vector Machines* - SVM) [40] por ser adequado a problemas com poucas amostras de treinamento e por ser bastante utilizado nos trabalhos relacionados [15].

3.3.1 Máquina de Vetores de Suporte

Segundo Vapnik [40], a função de decisão mais adequada é aquela para qual a distância entre os conjuntos das amostras de treinamento é maximizada. Utilizando este princípio, o SVM coloca a superfície de decisão exatamente entre o limite das duas classes, reduzindo a probabilidade de erro de classificação e maximizando a capacidade de generalização do classificador.

Considerando um classificador binário, com dados de treinamento $x_i (i = 1, ..., m)$, possuindo classes correspondentes $y_i = \pm 1$, a função de decisão pode ser formulada como na equação (3.12), onde $\mathbf{w}^T\mathbf{x} + b = 0$ representa o hiperplano ótimo de separação, b o viés e $\mathbf{w}$ o vetor de pesos ortogonal ao hiperplano separador. Os valores de $(\mathbf{w}, b)$ podem ser obtidos resolvendo um problema de otimização, em que se busca maximizar a margem (ρ) entre os hiperplanos de suporte, implicando em maximizar $\frac{2}{\|w\|}$ ou minimizar $\frac{1}{2}\|\mathbf{w}\|^2$ com as restrições lineares $y_i(\mathbf{w}^T\mathbf{x_i} + b) \geq 1, 1 \leq \mathrm{i} \leq \mathrm{m}$.

$$g(\mathbf{x}) = sign(\mathbf{w}^T\mathbf{x} + b) \tag{3.12}$$

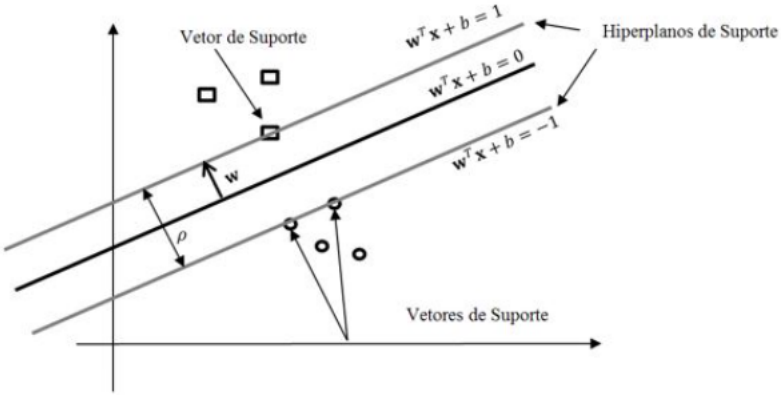


Figura 3.11: O hiperplano ótimo de separação com margem máxima ρ. Os vetores de suporte são as amostras que satisfazem as equações $g(\mathbf{x}) = 1$ ou $g(\mathbf{x}) = -1$. Adaptado de Hammel [3].

Truque do Kernel

Apesar de originalmente ser formulado como um classificador linear, a SVM pode lidar com problemas não lineares. Para isso, pode-se utilizar uma função, chamada função de *kernel* ou truque do *kernel*, que possibilita que o espaço original seja mapeado em um espaço de produto escalar de alta-dimensão onde os dados possivelmente podem ser linearmente separáveis. Na Figura 3.12 é possível visualizar um exemplo de mapeamento de um espaço bidimensional para um espaço tridimensional em que o dados de exemplo se tornam linearmente separáveis.

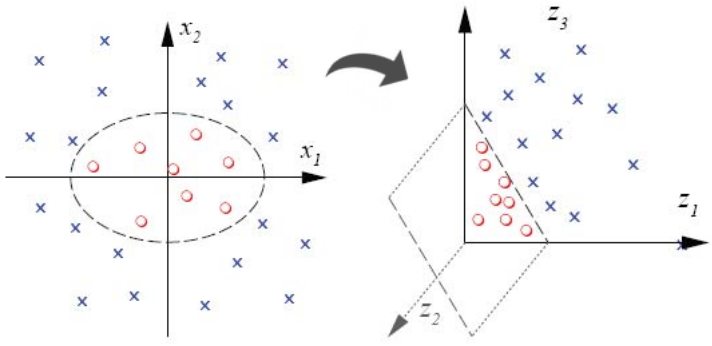


Figura 3.12: Mapeamento de dados para um espaço de características de maior dimensão.

Utilizando uma função vetorial não-linear $h(x, x')$, que mapeia um vetor de entrada n-dimensional para um espaço de características l-dimensional, a função de decisão linear no

espaço de características é dada pela equação (3.13) onde $\mathbf{w}$ é um vetor l-dimensional e b é o termo independente. .

$$g(\mathbf{x}) = \mathbf{w}^T h(\mathbf{x}, \mathbf{x}') + b \tag{3.13}$$

A escolha da função de *kernel* e os valores apropriados dos seus parâmetros podem afetar consideravelmente o desempenho do classificador SVM. Nesse trabalho, foi escolhido o *kernel* RBF por possuir somente um parâmetro livre (γ). Segundo Hsu *et al.* [48], o *kernel* RBF é uma boa escolha inicial, uma vez que o *kernel* linear é apontado como um caso especial da função RBF [49] e o kernel sigmoidal pode ter comportamento semelhante ao RBF para certos parâmetros [50].

Segundo Taylor e Cristianini [51], o parâmetro γ controla a flexibilidade da função de *kernel* e é equivalente a $\gamma = \frac{1}{2\sigma^2}$, em que σ controla o espalhamento da gaussiana. Valores pequenos de γ permitem o classificador ajustar todos os rótulos havendo risco de sobre ajustamento (*overfitting*). A equação (3.14) explicita o *kernel* RBF.

$$\begin{aligned} k(x, x') &= exp\left(-\gamma \left\|x - x'\right\|^2\right) \\ &= exp\left(-\frac{\left\|x - x'\right\|_2^2}{\frac{1}{2\sigma^2}}\right) \end{aligned} \tag{3.14}$$

3.3.2 Seleção de Parâmetro de Kernel

A seleção dos parâmetros do *kernel* do SVM é crítica para obter um bom desempenho. Inicialmente, Vapnik [40] recomenda a escolha manual dos parâmetros do *kernel* pelo especialista baseado no conhecimento *a priori* do conjunto de dados a ser avaliado. No entanto, o processo de seleção de parâmetros de maneira empírica pode resultar em um acurácia inferior. Neste trabalho, para selecionar o parâmetros do *kernel* RBF de forma automática é utilizada a Busca em Grade [52].

No contexto de aprendizado de máquina, a Busca em Grade se refere ao processo de busca exaustiva sobre um subconjunto do espaço de trabalho. Sendo assim, para cada abordagem de seleção de característica proposta, foi realiza uma a busca exaustiva visando encontrar o valor de γ que maximiza a acurácia do classificador. O processo de busca exaustiva é bastante simples de ser programado e apresenta bons resultados na busca dos melhores

parâmetros. Sua desvantagem é o tempo de processamento, já que faz uma busca linear no espaço de parâmetros. Outra questão fundamental no método de Busca em Grade é a delimitação do espaço a ser investigado quando não possuímos um conhecimento prévio dos dados a serem classificado. A busca em um espaço muito amplo resulta em um tempo de processamento alto e muitas vezes com resultados pouco efetivos. Neste trabalho, a busca pelos valores de gamma foi delimitada de 0 a 100, já que realizando testes aleatórios com valores maiores e menores do que este intervalo não foram observas melhoras significativas na acurácia.

Capítulo 4

Avaliação Experimental

Nesse capítulo são apresentados e discutidos os experimentos realizados durante a pesquisa. Inicialmente, na Seção 4.1, as ferramentas e o cenário utilizado são apresentados. Em seguida, na Seção 4.2, são apresentados os resultados obtidos tanto com as Distâncias Empíricas quanto com o método das distâncias obtidas através da seleção de características por correlação CFS. Por fim, na Seção 4.3, é realizada uma comparação dos resultados obtidos com outros trabalhos encontrados.

4.1 Condução dos Experimentos

Para verificar a metodologia apresentada no Capítulo 3, foi elaborada uma aplicação utilizando a linguagem de programação C++ e a biblioteca de visão computacional OpenCV [53] na sua versão 2.4.X. Para o classificador SVM, a implementação utilizada foi a da biblioteca LibSVM [54]. A implementação da seleção CFS utilizada foi a disponível no Weka [46] realizada por Hall [45], autor do método.

Conforme demonstrado por alguns autores [32], [17] e [25], os métodos de reconhecimento de expressões faciais que se baseiam nas características geométricas da face possuem performance similar ou melhor que os modelos baseados na textura da face. No entanto, esses os métodos necessitam de uma detecção e rastreamento facial precisa e confiável, impactando diretamente no desempenho final do sistema reconhecedor.

Sabendo disso, todos os experimentos foram conduzidos utilizando dois PDMs. O primeiro obtém os pontos de controle através de um AAM que é fornecido juntamente com

a base de dados Cohn-kanade [2]. É esperado que os experimentos conduzidos com este PDM encontrem melhor desempenho que os demais, já que o AAM foi montado utilizando os exemplos da mesma base de dados (Cohn-kanade [2]) utilizada na fase de testes e, portanto, é chamada de abordagem dependente de indivíduo (*person-dependent*). Já o segundo, foi obtido através de uma abordagem CLM [42] treinado utilizando a base de dados MUCT [4], que contém pessoas de múltiplas etnias e diversas idades em diferentes condições de iluminação. Sendo assim, é esperado que os resultados obtidos através desse PDM "mais genérico"ou independente de indivíduo (*person-independent*) sejam piores que o anterior, já que não é especializado na base de dados utilizada porém é mais adequado para situações reais.

Para verificar o desempenho das duas abordagens de seleção de características geométricas propostas, foi utilizada a base de dados Cohn-kanade Estendida (Ck+). Ela contém 309 sequências de imagens de 123 estudantes universitários (de 18 a 30 anos de idade) manifestando as emoções básicas descritas na Tabela 1.1 mostrada no Capítulo 1. Cada sequência de imagens inicia com o indivíduo manifestando a expressão neutra, totalizando 309 exemplos, e evolui até o auge da manifestação expressão. Dessas 309 sequências imagens, 45 delas apresentam exemplos de raiva, 59 de aversão, 25 de medo, 69 de felicidade, 28 de tristeza e 83 de surpresa. Vale lembrar que, além das emoções básicas, a base de dados Ck+ possui ainda 18 exemplos de sequências de imagens que demonstram a expressão de desprezo, no entanto a expressão foi desconsiderada desse trabalho por possuir poucos exemplos desta expressão.

Para cada método de seleção de características proposto, foram realizados experimentos tanto incluindo o reconhecimento da expressão Neutra (problema de 7 classes) quanto eliminando a detecção da expressão Neutra (problema de 6 classes) do classificador já que, após a seleção dos exemplos de imagens rotuladas com a expressão facial manifestada, a quantidade de exemplos da expressão neutra é maior que a das demais classes o que torna o classificador enviesado. É importante destacar que, para lidar com o problema do desbalanceamento, foram realizados experimentos inicias utilizando uma abordagem baseada em custos de classificação diferentes para as classes, no entanto, não foi possível definir o custo de cada classe que obtivesse ganho de acurácia do classificador.

4.2 Experimentos

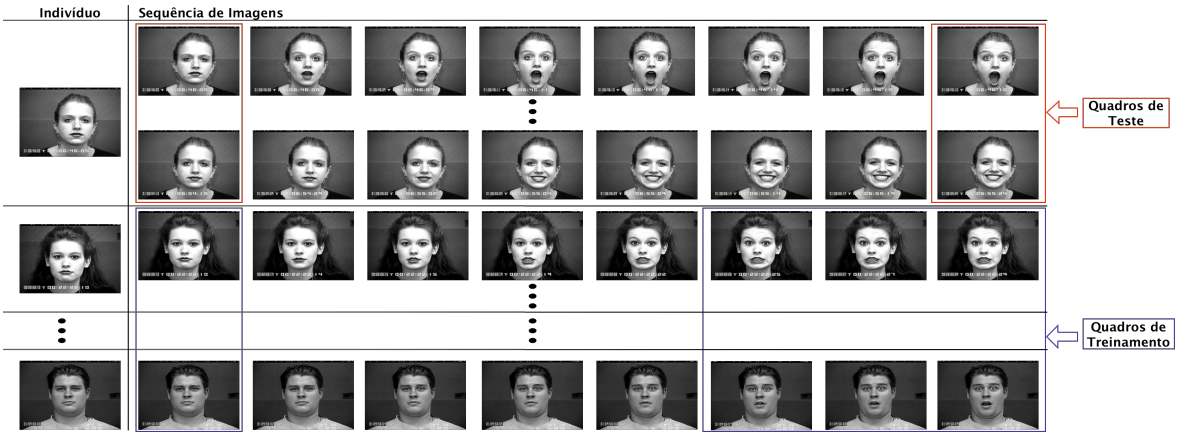


Figura 4.1: Uma possível dobra para o protocolo de amostragem utilizado.

Neste trabalho, para avaliar a performance do classificador, foi utilizada a estratégia de amostragem ilustrada na Figura 4.1, em uma abordagem de validação cruzada com o método *k-fold*, em que $k = 10$, para avaliar a capacidade de generalização do classificador. Nela, a separação do conjunto de testes em cada dobra é realizada baseando-se na sequência de imagens em que, para cada separação de dobra do *k-fold*, separa-se as sequências de imagens que farão parte do conjunto de treinamento das que farão parte do conjunto de testes. Com o objetivo de aumentar a quantidade de exemplos de cada expressão, para o grupo de treinamento, são selecionadas a primeira (quando considerada a expressão Neutra) e as três últimas imagens. Já para o grupo de teste, são consideradas somente a primeira (quando considerada a expressão Neutra) e a última imagem de cada sequência. Esta estratégia de amostragem, ou similar, foi adotada na maioria dos trabalhos comparados na Seção 4.3. Os resultados obtidos para as duas abordagens de seleção apresentadas no Capítulo 3, Distâncias Empíricas e Distâncias CFS, são apresentados a seguir.

Distâncias Empíricas

Para o problema de 6 classes, utilizando a abordagem dependente de indivíduo, a abordagem das *Distâncias Empíricas* obteve média $77,66\%$ de taxa de reconhecimento. No entanto, com a redução da qualidade da detecção dos pontos de controle, quando utilizado o método PDM independente de indivíduo, a taxa de reconhecimento obtida diminuiu para 74% em média.

Tabela 4.1: Acurácia média e desvio padrão por classe obtidos com a estratégia de seleção de atributos das *Distâncias Empíricas* aplicada ao problema de 6 classes.

	Raiva	Aversão	Medo	Felicidade	Tristeza	Surpresa
PDM Dep.	$\mu = 61{,}66\%$ $\sigma = 42{,}52$	$\mu = 72{,}86\%$ $\sigma = 20{,}10$	$\mu = 62{,}5\%$ $\sigma = 31{,}73$	$\mu = 88{,}60\%$ $\sigma = 16{,}50$	$\mu = 55{,}74\%$ $\sigma = 40{,}62$	$\mu = 94{,}36\%$ $\sigma = 10{,}00$
PDM Indep.	$\mu = 58{,}61\%$ $\sigma = 37{,}13$	$\mu = 75{,}94\%$ $\sigma = 20{,}52$	$\mu = 27{,}5\%$ $\sigma = 34{,}25$	$\mu = 85{,}71\%$ $\sigma = 17{,}81$	$\mu = 50{,}18\%$ $\sigma = 28{,}63$	$\mu = 89{,}92\%$ $\sigma = 10{,}23$

Tabela 4.2: Acurácia média e desvio padrão por classe obtida adicionando a expressão Neutra ao reconhecimento da expressões utilizando a abordagem das *Distâncias Empíricas*.

	Neutra	Raiva	Aversão	Medo	Felicidade	Tristeza	Surpresa
PDM Dep.	$\mu = 89\%$ $\sigma = 10{,}30$	$\mu = 8{,}88\%$ $\sigma = 22{,}86$	$\mu = 64{,}36\%$ $\sigma = 21{,}93$	$\mu = 55\%$ $\sigma = 28{,}38$	$\mu = 93{,}03\%$ $\sigma = 9{,}97$	$\mu = 0\%$ $\sigma = 0$	$\mu = 93{,}25\%$ $\sigma = 9{,}92$
PDM Indep.	$\mu = 92{,}33\%$ $\sigma = 6{,}67$	$\mu = 12{,}27\%$ $\sigma = 17{,}00$	$\mu = 66{,}16\%$ $\sigma = 21{,}02$	$\mu = 22{,}5\%$ $\sigma = 34{,}25$	$\mu = 81{,}78\%$ $\sigma = 18{,}93$	$\mu = 9{,}62\%$ $\sigma = 14{,}94$	$\mu = 90{,}71\%$ $\sigma = 9{,}46$

O desvio padrão foi de $9,55$ e $11,13$ respectivamente. A Tabela 4.1 mostra a acurácia média e o desvio padrão por classe após 10 iterações.

No problema de 7 classes, a taxa de reconhecimento média foi mantida em $77,33\%$ com desvio padrão de $7,03$ ao utilizar o alinhamento do PDM dependente de indivíduo. Para o método independente de indivíduo, a taxa de reconhecimento média obtida foi de 77% com desvio padrão de $7,40$. A Tabela 4.2 mostra a acurácia média por classe obtida após a realização das 10 iterações do *k-fold* quando incluída a expressão neutra.

Analisando as Tabelas 4.1 e 4.2, podemos observar que a adição da expressão Neutra, em maior quantidade, afeta o desempenho do classificador principalmente para as classes a Tristeza e a Raiva. Além disso, nota-se que efetivamente o método de seleção das *Distâncias Empíricas* somente obtém bons resultados considerando a detecção das expressões Neutra,

Tabela 4.3: Acurácia média por classe obtida com a seleção de características das *Distâncias CFS* no problema de 6 classes.

	Raiva	Aversão	Medo	Felicidade	Tristeza	Surpresa
PDM Dep.	$\mu = 91{,}05\%$ $\sigma = 12{,}73$	$\mu = 86{,}47\%$ $\sigma = 15{,}82$	$\mu = 77{,}5\%$ $\sigma = 24{,}86$	$\mu = 97{,}14\%$ $\sigma = 6{,}02$	$\mu = 88{,}88\%$ $\sigma = 22{,}04$	$\mu = 94{,}36\%$ $\sigma = 10{,}00$
PDM Indep.	$\mu = 63{,}83\%$ $\sigma = 25{,}21$	$\mu = 77{,}02\%$ $\sigma = 24{,}46$	$\mu = 57{,}5\%$ $\sigma = 40{,}90$	$\mu = 88{,}74\%$ $\sigma = 13{,}09$	$\mu = 51{,}66\%$ $\sigma = 28{,}57$	$\mu = 93{,}35\%$ $\sigma = 9{,}87$

Felicidade e Surpresa.

Distâncias CFS

Utilizando a abordagem das *Distâncias CFS* para o problemas de 6 classes e o PDM dependente de indivíduo, foi obtida a taxa de reconhecimento média de $91,33\%$. No entanto, ao utilizar o PDM independente de indivíduo, a taxa de reconhecimento média obtida foi de $78,33\%$, melhorando os resultados obtido com o método das *Distâncias Empíricas*. O desvio padrão foi de $16,22$ e $12,19$ respectivamente. Na Tabela 4.3 pode-se visualizar a acurácia média por classe obtida após 10 iterações utilizando o método das *Distâncias CFS* aplicado ao problema de 6 classes.

Além de detectar as seis expressão básicas, a abordagem das *Distâncias CFS* também reconhece a expressão neutra. No entanto, neste caso a taxa de reconhecimento média obtida com o PDM dependente de indivíduo foi de $88,66\%$ com desvio padrão de $13,51$. Porém, ao utilizar o método independente de indivíduo, a taxa de reconhecimento a taxa de reconhecimento atingida foi de $81,16\%$ com desvio padrão de $11,95$. Na Tabela 4.4 pode-se visualizar a acurácia média por classe obtida após 10 iterações do *k-fold* ao utilizar o método das *Distâncias CFS* aplicado ao problema de 7 classes.

Analisando os resultados por classe, apresentados nas Tabelas 4.3 e 4.4, verificamos que as Distâncias CFS conseguem detectar todas as expressões satisfatoriamente, apesar do alto desvio padrão em algumas classes como o Medo. Possivelmente, o maior número de distâncias consideradas na abordagem das *Distâncias CFS* auxiliou nos melhores resultados

Tabela 4.4: Acurácia média por classe obtida com a seleção de características das *Distâncias CFS* no problema de 7 classes.

	Neutra	Raiva	Aversão	Medo	Felicidade	Tristeza	Surpresa
PDM Dep.	$\mu = 94{,}66\%$ $\sigma = 4{,}21$	$\mu = 63\%$ $\sigma = 21{,}26$	$\mu = 85{,}83\%$ $\sigma = 18{,}34$	$\mu = 75\%$ $\sigma = 26{,}35$	$\mu = 95{,}89\%$ $\sigma = 6{,}63$	$\mu = 52{,}96\%$ $\sigma = 36{,}83$	$\mu = 94{,}36\%$ $\sigma = 10{,}00$
PDM Indep.	$\mu = 94{,}33\%$ $\sigma = 8{,}47$	$\mu = 37{,}94\%$ $\sigma =30{,}16$	$\mu = 60{,}02\%$ $\sigma = 24{,}10$	$\mu = 47{,}5\%$ $\sigma = 41{,}58$	$\mu = 88{,}74\%$ $\sigma = 13{,}09$	$\mu = 16{,}65\%$ $\sigma = 20{,}41$	$\mu = 91{,}82\%$ $\sigma = 13{,}71$

quando comparados a abordagem das *Distâncias Empíricas*.

4.3 Comparação e Discussão

Ao analisar os resultados obtidos, nota-se que a abordagem baseada em *Distâncias CFS* superou à baseada em *Distâncias Empíricas*. Por outro lado, a abordagem das *Distâncias Empíricas* possui a vantagem das distâncias consideradas já estarem definidas independente da base de dados, enquanto que a abordagem das *Distâncias CFS* tem de avaliar uma base de dados para obter as distâncias mais adequadas. Como esperado, o método dependente de indivíduo obteve melhores resultados em relação a abordagem independente de indivíduo.

Por ter sido uma das primeiras bases de dados disponíveis contendo informações da expressão facial manifestada, a base de dados Cohn-kanade [2] naturalmente atraiu bastante atenção dos grupos de pesquisa da área. Alguns trabalhos que utilizam características geométricas, imagens estáticas e SVM para classificação de padrões, estão listados na tabela 4.5. Como pode-se observar, a taxa de reconhecimento dos métodos propostos estão de acordo com as taxas encontradas trabalhos de outros autores.

Neste trabalho foi obtido $91,33\%$ de acurácia utilizando 44 atributos selecionados através da abordagem de seleção de características da *Distâncias CFS*, além de $78,33\%$ de taxa de reconhecimento utilizando a abordagem das *Distâncias Empíricas*, ambas utilizando SVM com *kernel* RBF para a etapa de classificação de características e considerando o PDM dependente de indivíduo. Embora o sistema proposto por Kotsia *et al.* [34] tenha obtido melhor

desempenho, chegando a $99,7\%$, nele a inicialização dos pontos de controle se dá através de um processo manual, não sendo totalmente automatizado em contraste com o método proposto que é totalmente automatizado, além do número de pontos de controle utilizado ser maior o que pode prejudicar a execução em tempo real já que o quanto maior for o número de pontos controle do modelo maior é a necessidade o processamento.

A fim de melhorar o desempenho do seu sistema, Hsieh *et al.*[31] incluiu algumas características de aparência e desconsiderou a expressão de Tristeza. Chen *et al.* [33] também obteve bons resultados mas não considerou a expressão Neutra nos seus experimentos. Podemos concluir que o método proposto das *Distâncias CFS* alcançou resultados compatíveis com os encontrados em outros trabalhos dessa área de pesquisa como os de Xiao *et al.* [55] e Shan *et al.* [17], em experimentos utilizando a base de dados Cohn-kande[2], demonstrando a contribuição deste trabalho. A Tabela 4.5 mostra uma comparação de performance obtida em trabalhos da área.

Tabela 4.5: Comparação de performance das duas abordagens propostas nesse trabalho utilizando o PDM independente de indivíduo com métodos de reconhecimento de expressões faciais que utilizaram a Cohn-kanade [2] nos experimentos.

Método	**Ano**	**N Classes**	**Taxa de Reconhecimento**
Barlett et al. [56]	2005	7 classes	90,9%
***Kotsia et al. [34]	2007	6 classes	99,7
Shan et al. [17]	2009	7 classes 6 classes	91,40% 95,10%
Tsai et al. [57]	2010	7 classes	98,59%
Xiao et al. [55]	2011	6 classes	96,57%
Zhang et al. [58]	2011	6 classes	94,48%
Lajevardi and Hussain [59]	2012	7 classes	91,9%
Chen et al. [33]	2012	**7 classes	95%
Hsu et al. [60]	2014	7 classes	89,6%
Saeed et al. [29]	2014 2014	6 classes 7 classes	83,01% 73,06%
Hsieh et al. [31]	2015	*6 classes	94,7%
Distâncias Empíricas	2016	7 classes 6 classes	77,33% 77,66%
Distâncias CFS	2016	7 classes 6 classes	88,66% 91,33%

*A expressão Tristeza foi desconsiderada.

**A expressão Desprezo foi considerada e a expressão Neutra foi desconsiderada.

***O método não é totalmente automatizado, sendo necessária marcação manual.

Capítulo 5

Conclusão

As expressões faciais são formas de comunicação não verbal amplamente utilizadas em nosso cotidiano. Elas externam, consciente ou inconscientemente, nossas manifestações em relação aos estímulos internos e externos a que somos submetidos. Buscando melhorar a interação homem-máquina, o trabalho propôs inferir automaticamente a emoção expressa na face de um usuário de computador, baseando-se nas informações da geometria facial humana. Frente a isso, durante o levantamento bibliográfico, foram pesquisados diferentes métodos de extração e seleção de características faciais que possibilitam a classificação da expressão facial humana dentre as emoções básicas de Ekman [1].

Inicialmente, para rastrear o movimento dos músculos faciais foram estudadas técnicas de Modelos Pontuais de Distribuição para garantir a qualidade da representação computacional do estado da face. Em seguida, durante a etapa de extração e seleção de características, dois métodos de seleção de atributos foram propostos: *Distâncias Empíricas* e *Distâncias CFS*. O primeiro foi concebido baseando-se na experiência empírica dos usuários ao analisar o posicionamento de pontos de controle que descrevem o formato da face manifestando uma expressão. O segundo, aplicou o algoritmo de seleção de características por correlação, escolhendo o subconjunto de atributos mais correlacionados com as expressões faciais (classes) porém não-correlacionados entre si. Por último, para a classificação de padrões foi escolhido SVM.

Sabendo que os métodos que se baseiam na geometria facial dependem diretamente da qualidade do mecanismo rastreador da face, para analisar o desempenho das abordagens propostas, foram utilizados tanto um Modelo Pontual de Distribuição construído com diversas

imagens de exemplo da face humana, possuindo diferentes condições de iluminação, etnia e gênero, quanto um específico para a base de dados utilizada para avaliar o experimento, a Cohn-kanade [2]. Os resultados obtidos mostraram que ambas produziram resultados relevantes, que são compatíveis com o atual estado da arte da área de pesquisa.

Em trabalhos futuros, outras técnicas de aprendizado de máquina podem ser avaliadas, em especial as baseadas em aprendizagem profunda que já mostraram ser promissores nesta área de pesquisa como demonstrado em [61]. A fim de melhorar a performance dos métodos propostos, abordagens de balanceamento baseado em custos devem ser investigadas para tentar minimizar o impactado na performance final dos classificadores. Além disso, mais experimentos podem ser executados utilizando outras bases de dados para verificar a performance dos métodos propostos em um cenário mais amplo.

REFERÊNCIAS

[1] P. Ekman, "Basic emotions," in *The Handbook of Cognition and Emotion*, T. Dalgleish and T. Power, Eds. John Wiley & Sons, Ltd, 1999, ch. 3, pp. 45–60. [Online]. Available: http://dx.doi.org/10.1002/0470013494.ch3

[2] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010, pp. 94–101.

[3] L. H. Hamel, *Knowledge Discovery with Support Vector Machines*. Wiley-Interscience, 2011, vol. 3.

[4] S. Milborrow, J. Morkel, and F. Nicolls, "The MUCT Landmarked Face Database," *Pattern Recognition Association of South Africa*, vol. 201, no. 0, 2010.

[5] D. Baggio, *Mastering OpenCV with Practical Computer Vision Projects*. Packt Publishing, Limited, 2012.

[6] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models - their training and application," *Computer vision and image understanding*, vol. 61, pp. 38–59, 1995.

[7] C. Darwin, *The Expression of the Emotions in Man and Animals*, 1872, the original was published 1898 by Appleton, New York. Reprinted 1965 by the University of Chicago Press, Chicago and London,.

[8] A. Ortony and T. J. Turner, "What's basic about basic emotions?" *Psychol*

Rev, vol. 97, no. 3, pp. 315–331, Jul. 1990. [Online]. Available: http://view.ncbi.nlm.nih.gov/pubmed/1669960

[9] P. Ekman and W. V. Friesen, *Pictures of Facial Affect*. Consulting Psychologists Press, 1976.

[10] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *2005 IEEE International Conference on Multimedia and Expo*, July 2005, pp. 5 pp.–.

[11] M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, and J. Budynek, "The japanese female facial expression (jaffe) database," in *Proceedings of third international conference on automatic face and gesture recognition*, 1998, pp. 14–16.

[12] P. Ekman, W. Friesen, and J. Hager, *Facial Action Coding System (FACS): Manual*. Salt Lake City (USA): A Human Face, 2002.

[13] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," *Pattern recognition*, vol. 36, no. 1, pp. 259–275, 2003.

[14] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 1, pp. 39–58, 2009.

[15] A. Danelakis, T. Theoharis, and I. Pratikakis, "A survey on facial expression recognition in 3d video sequences," *Multimedia Tools and Applications*, vol. 74, no. 15, pp. 5577–5615, 2015.

[16] P. Liu, S. Han, Z. Meng, and Y. Tong, "Facial expression recognition via a boosted deep belief network," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014, pp. 1805–1812.

[17] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image Vision Comput.*, vol. 27, no. 6, pp. 803–816, May 2009. [Online]. Available: http://dx.doi.org/10.1016/j.imavis.2008.08.005

[18] Y.-H. Byeon and K.-C. Kwak, "Facial expression recognition using 3d convolutional neural network," *International Journal of Advanced Computer Science and Applications*, vol. 5, no. 12, 2014.

[19] Y.-I. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, no. 2, pp. 97–115, 2001.

[20] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1424–1445, 2000.

[21] T. Wu, S. Fu, and G. Yang, *Advances in Brain Inspired Cognitive Systems: 5th International Conference, BICS 2012, Shenyang, China, July 11-14, 2012. Proceedings*. Springer Berlin Heidelberg, 2012, ch. Survey of the Facial Expression Recognition Research, pp. 392–402.

[22] V. Bettadapura, "Face expression recognition and analysis: The state of the art," *arXiv preprint arXiv:1203.6722*, pp. 1–27, 2012.

[23] C. Sumathi, T. Santhanam, and M. Mahadevi, "Automatic facial expression analysis a survey," *International Journal of Computer Science and Engineering Survey*, vol. 3, no. 6, p. 47, 2012.

[24] S. Rifai, Y. Bengio, A. Courville, P. Vincent, and M. Mirza, "Disentangling factors of variation for facial expression recognition," in *European Conference on Computer Vision*. Springer, 2012, pp. 808–822.

[25] M. F. Valstar and M. Pantic, "Combined support vector machines and hidden markov models for modeling facial action temporal dynamics," in *Proceedings of the 2007 IEEE International Conference on Human-computer Interaction*, ser. HCI'07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 118–127.

[26] S. Bashyal and G. K. Venayagamoorthy, "Recognition of facial expressions using gabor wavelets and learning vector quantization," *Engineering Applications of Artificial Intelligence*, vol. 21, no. 7, pp. 1056–1064, 2008.

[27] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 6, 2007.

[28] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, and M. Bartlett, "The computer expression recognition toolbox (cert)," in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. IEEE, 2011, pp. 298–305.

[29] A. Saeed, A. Al-Hamadi, R. Niese, and M. Elzobi, "Frame-based facial expression recognition using geometrical features," *Advances in Human-Computer Interaction*, p. 4, 2014.

[30] F. Tang and B. Deng, "Facial expression recognition using aam and local facial features," in *Third International Conference on Natural Computation (ICNC 2007)*, vol. 2. IEEE, 2007, pp. 632–635.

[31] C.-C. Hsieh, M.-H. Hsih, M.-K. Jiang, Y.-M. Cheng, and E.-H. Liang, "Effective semantic features for facial expressions recognition using svm," *Multimedia Tools and Applications*, pp. 1–20, 2015.

[32] P. Michel and R. El Kaliouby, "Real time facial expression recognition in video using support vector machines," in *Proceedings of the 5th International Conference on Multimodal Interfaces*. ACM, 2003, pp. 258–264.

[33] J. Chen, D. Chen, Y. Gong, M. Yu, K. Zhang, and L. Wang, "Facial expression recognition using geometric and appearance features," in *Proceedings of the 4th International Conference on Internet Multimedia Computing and Service*, ser. ICIMCS '12. ACM, 2012, pp. 29–33.

[34] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 172–187, 2007.

[35] G. B. Dijksterhuis and J. C. Gower, "The interpretation of generalized procrustes analysis and allied methods," *Food Quality and Preference*, vol. 3, no. 2, pp. 67–87, 1992.

[36] T. F. Cootes, G. J. Edwards, C. J. Taylor *et al.*, "Active appearance models," vol. 23, no. 6, 2001, pp. 681–685.

[37] L. P. Chew, "Constrained delaunay triangulations," in *Proceedings of the Third Annual Symposium on Computational Geometry*, vol. 4, no. 1-4. Springer, 1989, pp. 97–108. [Online]. Available: http://doi.acm.org/10.1145/41958.41981

[38] I. Matthews and S. Baker, "Active appearance models revisited," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, 2004.

[39] D. Cristinacce and T. F. Cootes, "Feature detection and tracking with constrained local models." in *BMVC*, vol. 1, no. 2, 2006, p. 3.

[40] V. N. Vapnik and V. Vapnik, *Statistical learning theory*. Wiley New York, 1998, vol. 1.

[41] P. Viola and M. J. Jones, "Robust real-time face detection," vol. 57, no. 2. Springer, 2004, pp. 137–154.

[42] J. M. Saragih, S. Lucey, and J. F. Cohn, "Deformable model fitting by regularized landmark mean-shift," *International Journal of Computer Vision*, vol. 91, no. 2, pp. 200–215, 2011.

[43] K. Faceli, A. C. Lorena, J. Gama, and A. Carvalho, "Inteligência artificial: Uma abordagem de aprendizado de máquina," *Livros Técnicos e Científicos*, 2011.

[44] H. Soyel and H. Demirel, *Image Analysis and Recognition: 4th International Conference, ICIAR 2007, Montreal, Canada, August 22-24, 2007. Proceedings*. Springer Berlin Heidelberg, 2007, ch. Facial Expression Recognition Using 3D Facial Feature Distances, pp. 831–838.

[45] M. A. Hall, "Correlation-based feature selection for discrete and numeric class machine learning," in *Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000), Stanford University, Stanford, CA, USA, June 29 - July 2, 2000*. University of Waikato, Department of Computer Science, 2000, pp. 359–366.

[46] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," *SIGKDD Explor. Newsl.*, 2009.

[47] U. Fayyad and K. Irani, "Multi-interval discretization of continuous-valued attributes for classification learning," 1993.

[48] C.-W. Hsu, C.-C. Chang, C.-J. Lin *et al.*, "A practical guide to support vector classification," 2003.

[49] S. S. Keerthi and C.-J. Lin, "Asymptotic behaviors of support vector machines with gaussian kernel," *Neural computation*, vol. 15, no. 7, pp. 1667–1689, 2003.

[50] H.-T. Lin and C.-J. Lin, "A study on sigmoid kernels for svm and the training of non-psd kernels by smo-type methods," *submitted to Neural Computation*, pp. 1–32, 2003.

[51] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. New York, NY, USA: Cambridge University Press, 2004.

[52] S. Russell and P. Norvig, *Inteligência artificial*. Elsevier, 2004.

[53] G. Bradski *et al.*, "The opencv library," *Doctor Dobbs Journal*, vol. 25, no. 11, pp. 120–126, 2000.

[54] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.

[55] R. X. Q. Z. D. Z. P. Shi, "Facial expression recognition on multiple manifolds," *Pattern Recognition*, vol. 44, no. 1, pp. 107–116, 2011.

[56] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: machine learning and application to spontaneous behavior," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2. IEEE, 2005, pp. 568–573.

[57] H. H. Tsai, Y. S. Lai, and A. Y. C. Zhang, "Using svm to design facial expression recognition for shape and texture features," in *2010 International Conference on Machine Learning and Cybernetics*, vol. 5, 2010, pp. 2697–2704.

[58] L. Zhang and D. Tjondronegoro, "Facial expression recognition using facial movement features," *IEEE Transactions on Affective Computing*, vol. 2, no. 4, pp. 219–229, 2011.

[59] S. M. Lajevardi and Z. M. Hussain, "Automatic facial expression recognition: feature extraction and selection," *Signal, Image and Video Processing*, vol. 6, no. 1, pp. 159–169, 2012.

[60] F.-S. Hsu, W.-Y. Lin, and T.-W. Tsai, "Facial expression recognition using bag of distances," *Multimedia Tools and Applications*, 2014.

[61] A. Rao and N. Thiagarajan, "Recognizing facial expressions from videos using deep belief networks," Technical Report, Tech. Rep., 2009.

[62] W. A. Kass, M. and D. Terzopoulos, "Snakes: Active contour models," *INTERNATIONAL JOURNAL OF COMPUTER VISION*, 1988.

[63] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998, vol. 201.

[64] X. Bresson, S. Esedoglu, P. Vandergheynst, J.-P. Thiran, and S. Osher, "Fast global minimization of the active contour/snake model," *Journal of Mathematical Imaging and vision*, vol. 28, no. 2, 2007.

[65] A. L. Yuille, P. W. Hallinan, and D. S. Cohen, "Feature extraction from faces using deformable templates," *International journal of computer vision*, vol. 8, no. 2, 1992.

[66] A. Pentland and S. Sclaroff, "Closed-Form solutions for physically based shape modeling and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 7, pp. 715–729, 1991.

[67] D. Terzopoulos and D. Metaxas, "Dynamic 3d models with local and global deformations: Deformable superquadrics," pp. 606–615, 1990.

[68] B. G. Tabachnick and L. S. Fidell, *Using Multivariate Statistics (5th Edition)*, 5th ed. Allyn & Bacon, Mar. 2006.

[69] T. F. Cox and M. A. Cox, *Multidimensional Scaling*, ser. Chapman & Hall/CRC Monographs on Statistics & Applied Probab. CRC press, 2000.

[70] C. Brombin and L. Salmaso, "Multi-aspect permutation tests in shape analysis with small sample size." *Computational Statistics & Data Analysis*, vol. 53, no. 12, pp. 3921–3931, 2009.

[71] S. J. McKenna, S. Gong, R. P. Würtz, J. Tanner, and D. Banin, "Tracking facial feature points with gabor wavelets and shape models." in *International Conference on Audio-and Video-Based Biometric Person Authentication*, ser. Lecture Notes in Computer Science, J. Bigün, G. Chollet, and G. Borgefors, Eds. Springer, 1997, pp. 35–42.

[72] S. Xin and H. Ai, "Face alignment under various poses and expressions." in *International Conference on Affective Computing and Intelligent Interaction*, ser. Lecture Notes in Computer Science, J. Tao, T. Tan, and R. W. Picard, Eds. Springer, 2005, pp. 40–47.

[73] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in *Proceedings of the 10th European Conference on Computer Vision: Part IV*. Springer-Verlag, 2008, pp. 504–513.

[74] C. TECootes and A. Lanitis, "Active shape models: Evaluation of a multi-resolution method for improving image search," Citeseer, pp. 327–338, 1994.

Apêndice A

PDM - Modelos Pontuais de Distribuição

Os Modelos Pontuais de Distribuição (*PDM - Point Distribution Model*) buscam ajustar uma forma paramétrica pré-definida de uma estrutura, uma espécie de máscara deformável, à observação encontrada. A abordagem pode ser utilizada não somente para o rastreamento da face humana mas para a detecção e rastreamento de objetos em geral. Este método é assim conhecido pois possui como premissa que cada objeto ou estrutura seja representada através de um conjunto de pontos chamados de "pontos de controle" (ou pontos de referência, ou no inglês, "Landmarks"). Tais podem representar tanto as fronteiras quanto as características internas e externas de objetos. No caso da face, representa-se os contornos do rosto e dos elementos internos como olho, boca, nariz, lábios, sobrancelhas e etc.

A.1 Histórico

Os Modelos Pontuais de Distribuição fazem parte da família dos "Modelos Deformáveis". Estes são chamados assim pois se iniciam com uma configuração arbitrária, um contorno inicial, que se deforma até contornar e descrever a forma do objeto de interesse.

Este grupo de abordagens surgiu pela primeira vez em Kass et al.[62], sendo apresentado o Modelo de Contornos Ativos (ACM - Active Contour Models), os quais são um conjunto de pontos que se adaptam a uma estrutura a segmentar. O objetivo deste método é tentar ajustar uma curva, no caso uma *Spline* [63], sobre um objeto de uma imagem. Devido ao fato da *Spline* se mover constantemente buscando encontrar a borda do objeto e se ajustar aos seus contornos, a técnica é também conhecida como "Snakes". Surgiram ainda diversas

propostas de alterações no método original visando melhorar a sua performance como em [64].

Após o ACM, surgiram os Templates Deformáveis que utilizam formas semelhantes às que se pretendem destacar na imagem, como círculos, parábolas, etc., descritas por funções paramétricas. Yuille et al.[65], por exemplo, propõe um template deformável para detecção de um olho em imagens, representando a íris um círculo e as pálpebras com duas parábolas. Esta abordagem se mostrou desvantajosa pois a construção dos modelos é muito complexa e depende fortemente do objeto a considerar.

Além dos Templates Deformáveis, surgiram outros métodos que buscam modelar com maior precisão os objetos, como a Modelagem Física [66] e os Modelos Deformáveis Superquádricos [67]. No entanto, apesar dessas técnicas serem bastante intuitivas e explorarem bem as propriedades dos objetos, os modelos resultantes nem sempre produzem representações fiéis, podendo originar modelos que representam instâncias inválidas do objeto modelado.

Por último, os PDM extraem as principais características do objeto estudado através de técnicas estatísticas, por isso também são conhecidos como "Statistical Shape Models". Apresentado por Cootes et al.[6], os modelos pontuais são construídos a partir de um conjunto de exemplos do objeto. Cada exemplo então é manualmente descrito através de pontos, como por exemplo seu contorno e seus elementos internos. Estes pontos são colocados em locais equivalentes dentre o conjunto de exemplos. A técnica primeiramente propõe uma maneira automática de alinhar os pontos em localização equivalente para minimizar a variação da distância. Após isto, é realizada uma análise estatística por meio de uma Análise de Componentes Principais (PCA - Principal Component Analysis [68]), obtendo então um "Modelo Pontual de Distribuição". Através da média das posições de cada ponto de referência ao longo do conjunto de exemplos, e um número de parâmetros linearmente independentes que controlam as principais variações que cada ponto de referência pode sofrer obtido através da análise estatística, um modelo deformável pode ser ajustado a uma instância do objeto modelado.

Apesar de todos serem Modelos Deformáveis (ou flexíveis), assim como o PDM, nenhuma das abordagens anteriores realiza restrições globais no modelo, isto é, o modelo só pode ser deformado até que represente a deformação máxima do objeto. Além disso, um

Modelo Flexível deve realizar uma robusta interpretação mesmo em imagens com ruído, desordenadas ou com oclusão de partes do objeto de interesse. Sendo assim, a Seção A.2 descreve os passos para se obter um PDM.

A.2 Construindo um Modelo Pontual da Forma do Objeto

Nos PDM a estrutura de um objeto é representada através de um conjunto de n pontos, os quais podem estar em uma, duas ou três dimensões.

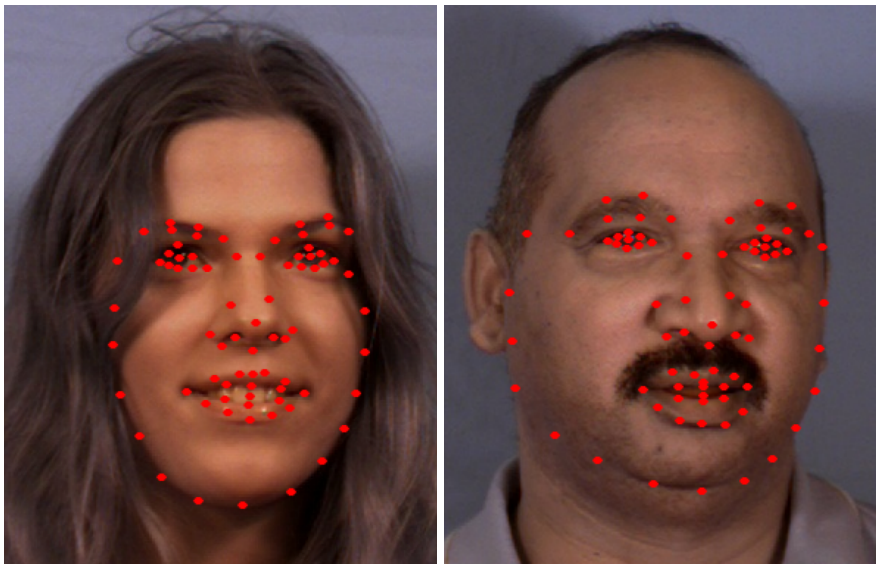
Podemos descrever o processo de obtenção do modelo nos seguintes passos:

1. Posicionar os pontos de referência em cada exemplo do conjunto de treinamento;
2. Alinhar os pontos equivalentes do conjunto de exemplos de treinamento;
3. Analisar as variações admissíveis do modelo;
4. Escolher o número de modos de variação para representar o objeto;

A.2.1 Posicionamento do pontos de Referência

Inicialmente, para que possamos construir um modelo deformável, devemos marcar manualmente os pontos de controle em imagens que contêm exemplos do mesmo objeto variando dentre todos os possíveis formatos que ele pode apresentar. Cootes *et al.* [6] recomenda que esse processo seja realizado manualmente já que não se conhece na literata um método que seja eficiente o suficiente na tarefa de obter a localização espacial correta de cada ponto de controle em diferentes imagens. O número n de pontos de controle que descreve o formato estrutural do objeto é arbitrário mas deve ser tal que permita descrever corretamente todos as formas e estruturas do objeto.

Assim, tomando a face humana como exemplo, devemos marcar pontos controle em diferentes imagens do mesmo objeto em posições equivalentes. A figura A.1 ilustra exemplos de marcação dos pontos.



(a) Exemplo de marcação manual de pontos.

(b) Exemplo de marcação dos pontos de controle em localização equivalente ao da imagem anterior.

Figura A.1: Exemplos de marcação dos prontos de controle em localização equivalente utilizando imagens da base de dados MUCT [4].

A.2.2 Alinhamento da Base de Treinamento

Possuindo exemplos do objeto marcados com um conjunto de pontos, devemos realizar uma análise estatística que possibilite descrever a variação da localização espacial de cada um dos pontos de controle. Para que isto seja possível, devemos então uniformizar os exemplos minimizando variações de rotação, translação e escala da forma encontrada, como podemos observar na figura A.2. Para solucionar o problema, faz-se necessário aplicar uma transformação em cada objeto de treinamento com intenção de alinhar todas as formas. O modelo original proposto por Cootes et al.[6] sugere a utilização da *"Análise Generalizada de Procrustes"*.

Suponha que um escalonamento tenha sido realizado em um conjunto de pontos por meio de dois métodos diferentes, dando origem a duas configurações distintas mas que representam o mesmo conjunto de objetos. A *Análise de Procrustes* dilata, translada, espelha e rotaciona uma das configurações para que os pontos se ajustem, da melhor maneira possível à outra, permitindo a comparação dos resultados. Segundo Dijksterhuis [35], este método é ideal para analisar dados oriundos de diferentes indivíduos.

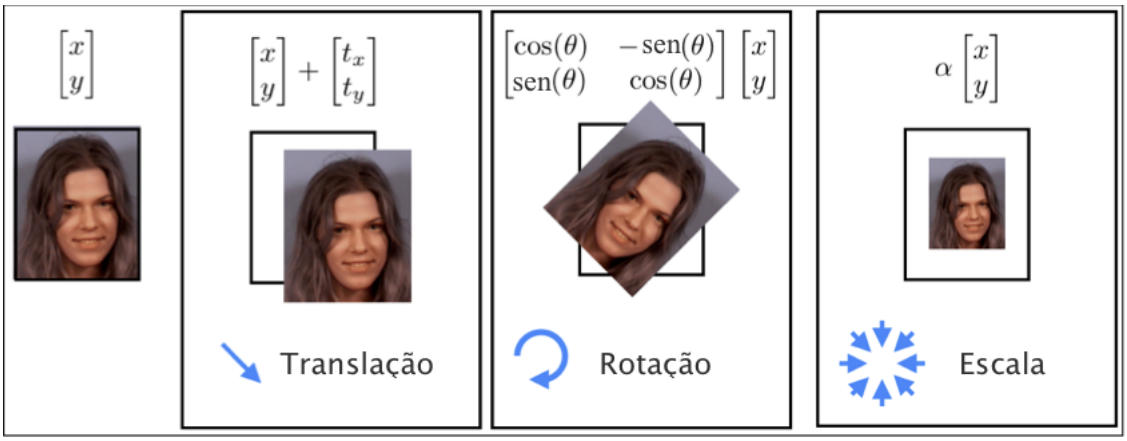


Figura A.2: Demonstração das variações de translação, rotação e escala. Adaptado de Baggio [5].

A solução teórica [69] supõe que uma configuração de n pontos em um espaço Euclidiano q-dimensional, com coordenadas dadas pela matriz $X1_{n_x q}$, precisa ser ajustada otimamente a outra configuração $X2$ dos mesmos n pontos em um espaço Euclidiano p-dimensional ($p \geq q$). Assume-se que o r-ésimo ponto na primeira configuração pode ser mapeado sobre o r-ésimo ponto da segunda.

Para comparar a forma de dois ou mais objetos, estes devem ser sobrepostos. Segundo Brombin *et al.* [70], a sobreposição ótima é encontrada calculando a melhor translação, rotação e escala (operações que podem ser observadas na figura A.2) aplicada aos objetos, em outras palavras, objetivo desta análise é aplicar uma transformação em cada objeto que minimize uma medida de diferença entre os formatos, chamada de distância de *Procrustes* obtida através da equação A.1. Essa medida nos informa a qualidade da sobreposição de dois objetos. Quanto mais próximos de zero mais o objeto $X2$ se aproxima do alinhamento ideal com o objeto $X1$.

$$P_d^2 = \sum_{j=1}^{n} [(x_{j1} - x_{j2})^2 + (y_{j1} - y_{j2})^2] \tag{A.1}$$

Por fim, para alinhar um conjunto de formas, ou seja, minimizar o efeito de rotação translação e escala dos objetos em relação a forma média, podemos seguir os seguintes passos:

Primeiramente devemos calcular a forma média ou *"Procrustes mean shape"* através da equação (A.2):

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{A.2}$$

onde N é o número de formas. Em seguida, devemos calcular o centro de massa de cada forma através da equação (A.3):

$$(\bar{x}, \bar{y}) = \left(\frac{1}{n}\sum_{j=1}^{n} x_j, \frac{1}{n}\sum_{j=1}^{n} y_j\right) \tag{A.3}$$

e então reposicionar a forma em relação ao seu centro de massa, transladando os pontos de forma que $(x, y) \rightarrow ((x - \bar{x}), (y - \bar{y}), ...)$.

Para remover a variação de escala dentre a base de dados de treinamento, podemos redimensionar os objetos de forma que o valor quadrático médio da distância a partir do ponto transladada para a origem seja 1. Sendo assim, temos:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (y_1 - \bar{y})^2 + ...}{k}} \tag{A.4}$$

onde k é o número de pontos da forma. A escala torna-se 1 quando as coordenadas dos pontos ajustadas da seguinte maneira:

$$(x, y) \rightarrow \left(\frac{(x - \bar{x})}{s}, \frac{(y - \bar{y})}{s}\right) \tag{A.5}$$

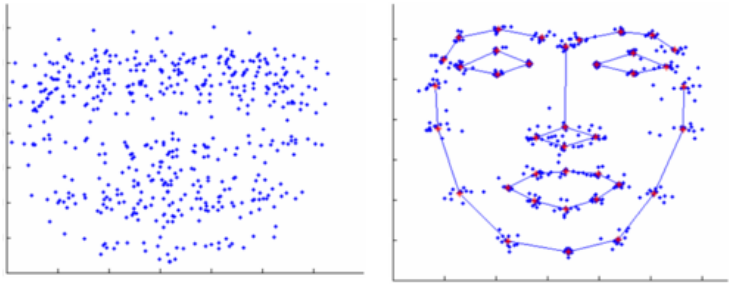
Por fim, para remover as diferenças de rotação, consideremos duas formas já alinhadas em relação a escala e translação. Seja $((x_1, y_1), ...)$ os pontos da forma 1 e $((w_1, z_1), ...)$ os pontos da forma 2. O angulo θ que melhor sobrepõe a forma 2 em relação a forma 1 pode ser calculada através da equação (A.6):

$$\theta = tan^{-}1\left(\frac{\sum_{i=1}^{k}(w_i y_i - z_i x_i)}{\sum_{i=1}^{k}(w_i x_i - z_i y_i)}\right) \tag{A.6}$$

A figura A.3 ilustra o resultado da aplicação da *Análise de Procrustes* nos pontos de controle de uma base de dados que contém imagens da face humana.

A.2.3 Estudo das Variações Admissíveis

A fim de encontrar as variações admissíveis do modelo, isto é, as maneiras segundo as quais os pontos de referência tendem a mover-se, devemos realizar uma análise estatística com o



(a) Dados plotados nas suas posições iniciais em 2D

(b) Formas alinhadas e forma média em vermelho

Figura A.3: Resultado da *Análise de Procrustes*

propósito de inferir a sua distribuição de probabilidade. Apesar de existirem vários métodos eficientes para este propósito, podemos simplificar o problema assumindo que os dados seguem uma distribuição Gaussiana e modelar os objetos com uma representação linear.

Assim, podemos aplicar uma clássica Análise de Componentes Principais (PCA - Principal Component Analysis) [68] a qual realiza uma transformação linear para um subespaço de menor dimensão que o original. Esta análise consiste em calcular os autovalores e autovetores da matriz de covariância S definida na equação A.7.

$$S = \frac{1}{N}\sum_{i=1}^{N_s}(x_i - \bar{x})(x_i - \bar{x})^T \tag{A.7}$$

Os modos de variação dos pontos de referência são descritos pelos autovetores de S, dados por p_i, tal que $Sp_i = \lambda_i p_i$ onde λ_i é o i^{esimo} autovalor de $S(\lambda_i \geq \lambda_{i+1})$ e $p_i^T p_i = 1$.

Após a aplicação do PCA, sabe-se que os autovalores da matriz de covariância S mais elevados descrevem a maior parte das variações admissíveis, e a proporção da variância total explicada por cada autovetor é igual ao autovalor correspondente. Assim, a maior parte da variância da forma pode ser explicada por um pequeno número de autovetores, que neste contexto, são chamados de modos de variação. Tal comportamento é ilustrado na figura A.2.3.

Dessa forma, cada objeto do conjunto de treinamento pode ser descrito pela forma média e pela combinação dos primeiros t autovetores como na equação A.8, onde $P = (p_1, p_2, ..., p_t)$ é a matriz dos primeiros t autovetores e $b = (b_1, b_2, ..., b_t)$ é o vetor de pesos de autovetor, isto é, o conjunto de parâmetros do modelo deformável.

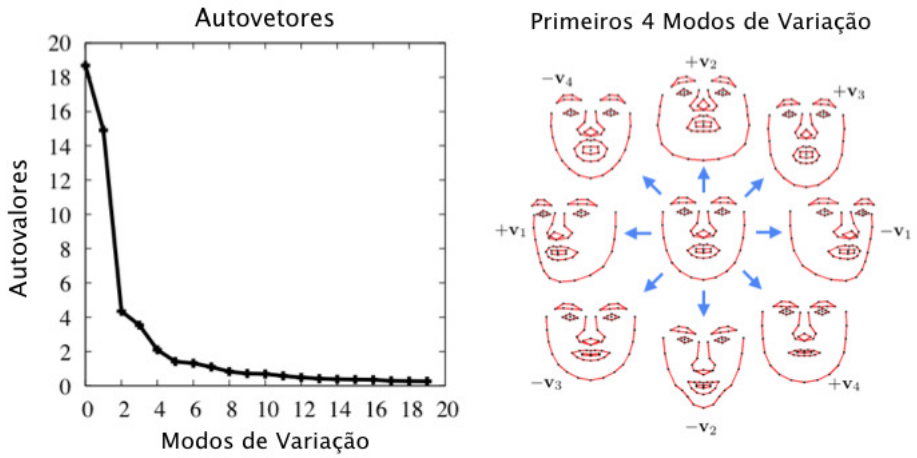


Figura A.4: Modos de Variação considerando o objeto face humana. Adaptado de Baggio [5].

$$x \cong \bar{x} + Pb \tag{A.8}$$

$$b = P^T(x - \bar{x}) \tag{A.9}$$

A.2.4 Escolha dos Modos de Variação Admissíveis

Esta etapa corresponde em escolher o número de autovetores (ou modos de variação), t, que representa a maior parte da variância total dos dados considerando que os termos residuais sejam considerados ruídos. Através do PCA, podemos criar novos objetos variando os parâmetros b_i dentro dos limites aceitáveis. Admitindo que os parâmetros b_i seguem uma distribuição Gaussiana, a maior parte das instâncias do objeto situa-se a menos de 3 desvios da média. Sabendo que a variância de b_i sobre o conjunto de treinamento é dada por λ_i, então a maioria das instâncias do objeto pode ser gerada estabelecendo um limite para b_i tal que:

$$-3\sqrt{\lambda_i} < b_i < 3\sqrt{\lambda_i} \tag{A.10}$$

Por se tratarem de modelos estatísticos que têm como insumo uma base de treinamento, é importante ressaltar que bases especializadas podem apresentar melhores resultados de ajuste do modelo de forma a uma observação do objeto, no entanto, em um cenário real, isso não é verdade. Para exemplificar, podemos considerar o objeto face humana; caso a base

de treinamento seja montada somente com imagens de uma pessoa específica, o subespaço que captura as variações da face é geralmente muito mais compacto do que quando montado com faces de múltiplas pessoas. Logo, neste exemplo, é desejável considerar uma base de treinamento com exemplos de face de diversas pessoas, possibilitando a generalidade do modelo.

A.3 Utilizando os PDMs em Problemas de Busca em Imagens

Em casos práticos onde necessitamos buscar a localização de cada um dos pontos de controle em uma imagem, como no caso do rastreamento dos músculos faciais, é necessário identificar o deslocamento de cada um dos pontos de controle. No entanto, já que em objetos não rígidos os pontos não se movem independentemente um do outro mas sim em sinergia, é necessário aplicar restrições aos deslocamentos estimados de acordo com o modelo PDM do objeto, de modo que somente formatos plausíveis sejam gerados para novos exemplos do objeto em imagens.

Ao longo dos anos surgiram diversas abordagens de busca pelo deslocamento dos pontos de controle que utilizam os princípios PDM de aplicar restrições de formato do objeto baseando-se em um modelo estatístico de deslocamento dos pontos, dentre as quais podemos citar: o *Modelo de Forma Ativa*, no inglês ASM - Active Shape Model, o *Modelo de Restrições Locais* ou CLM - Constraint Local Model e o *Modelo de Aparência Ativa* ou AAM - Active Appearance Model.

A.3.1 ASM - Modelo de Forma Ativa

O Modelo de Forma Ativa, ou simplesmente ASM, foi proposto inicialmente por Cootes *et al.* [6] sendo posteriormente aprimorado por diversos autores como em [71] [72] [73]. Nesta abordagem, primeiramente busca-se ao redor de cada ponto de controle o deslocamento mais adequado a nova configuração do objeto. Em seguida, cada deslocamento dos pontos de controle é então transformado em ajustes de formato e escala do modelo PDM, respeitando os limites dos formatos plausíveis.

Nesta abordagem de ASM, além de considerar a informação relativa à forma do objeto, é considerada também a informação ao redor de cada ponto de referência. Sendo assim, para a face humana, além de regular um modelo paramétrico ao formato do rosto, busca-se ajustar com maior precisão cada ponto que o compõe. Para isso, o ASM necessita montar um modelo de intensidade de níveis de cinza que representa cada ponto de referência. Vale ressaltar que, diferentemente do Modelo de Aparência Ativa descrito na Seção 3.1.1 ou do Modelo de Restrições Locais descrito na Seção 3.1.2 do Capítulo 3 onde o ajuste do modelo consiste em reproduzir o modelo de forma simultaneamente ao de aparência ou de fragmentos, as abordagens que seguem a ideia do ASM ajustam os pontos controle diretamente, buscando os deslocamento dos pixel na imagem de entrada.

Modelo de Perfil de Intensidade em torno dos pontos de controle

Sabendo que cada um dos pontos de controle equivalentes correspondem a mesma região do objeto, então, em diferentes instâncias do mesmo objeto, os níveis de cinza serão semelhantes. Sendo assim, o modelo de distribuição de pontos de um ASM possui também informação relativa às intensidades dos níveis de cinza em torno dos pontos de controle.

Uma forma de construir o modelo de intensidade de cada ponto de controle, é descrita a seguir:

- Para cada ponto de controle j da instância do objeto na imagem i do conjunto de treinamento, é extraído o perfil de intensidade g_{ij}, um vetor de dimensão n_p pixels (ou, em um caso mais genérico, uma matriz de busca n_{mn}), centrado nesse ponto.

- Extrai-se o perfil de intensidade do ponto de controle j na imagem i, tal que:

$$g_{ij} = \begin{bmatrix} g_{ij0} & g_{ij1} & \dots & g_{ijn_p-1} \end{bmatrix}^T, \tag{A.11}$$

  onde $g_{ijk} = I_i(y_k)$, sendo y_k o k^{esimo} ponto do perfil:

$$y_{ik} = P_{iInicio} + \frac{k+1}{n_p - 1}(P_{ifim} - P_{iinicio}), \tag{A.12}$$

  e $I_i(y_k)$ é o nível de cinza na imagem i neste ponto de referência.

  O perfil de intensidade a ser utilizado no momento de ajuste do modelo é obtido através

da média dos perfis de intensidade do ponto de referência dado por:

$$\bar{g}_{ij} = \frac{1}{N} \sum_{i=1}^{N} g_{ij} \tag{A.13}$$

Deste modo, podemos obter toda a informação sobre os perfis de intensidade necessária para a fase de ajuste do modelo. É importante ressaltar que o exemplo demonstrado utiliza apenas a informação relativa aos níveis de cinza para montar o perfil de intensidade. Outrossim, a fim de minimizar problemas como o de mudanças nas condições de iluminação ou diferenças de contraste, podem ser utilizadas diversas abordagens, como por exemplo, a da converter a imagem para escala logarítmica [5].

A figura A.3.1 exemplifica o modelo de perfil de intensidade montado a partir da base de dados de faces MUCT [4].

Figura A.5: Modelo de perfil de intensidade. Na imagem, cada ponto de referência está representado pelo seu perfil de intensidade médio.

Busca dos Deslocamentos dos Pontos de Controle

Possuindo o perfil de intensidade de cada ponto de controle, busca-se então determinar a localização de cada ponto em uma nova instância do objeto. Para isso, ao redor de cada ponto de controle anterior, procura-se pelo perfil de intensidade que mais se pareça com o perfil encontrado no modelo montado. Os perfis de pesquisa são comparados através de uma

função de semelhança, como por exemplo, a distância de *Mahalanobis*. Assim, seja g_{s_j} o perfil de intensidade da imagem e $\bar{g}_j$ a média dos perfis de intensidade para cada ponto de controle, denotando o sub-intervalo g_{s_j} centrado no d^{esimo} pixel de g_{s_j}, por $h(d)$, encontra-se o valor de d onde o sub-perfil $h(d)$ é mais semelhante a $\bar{g}_j$. Tal valor pode ser obtido minimizando a distância de *Mahalanobis* através da equação A.14.

$$f(d) = (h(d) - \bar{g}_j)^T S_{g_j}^{-1}(h(d) - \bar{g}_j)) \tag{A.14}$$

Correspondência entre o Modelo de Forma e os de Perfis de Intensidade

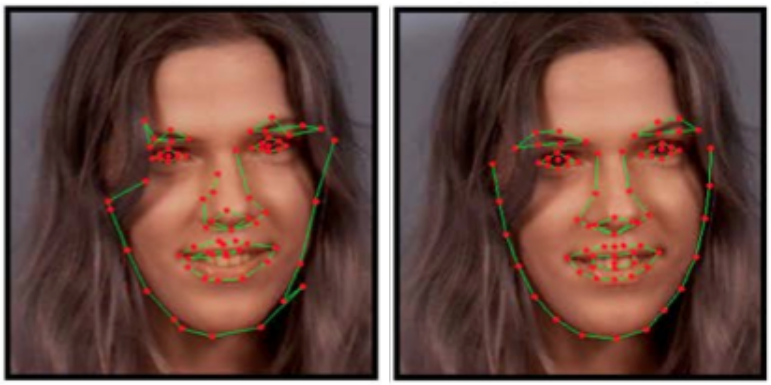
A partir da equação A.14, encontra-se a provável nova localização da característica e, consequentemente, o deslocamento dX necessário para ajustar o modelo. No entanto, é importante atribuir também as restrições impostas pelo modelo estatístico da forma pois, eventualmente, cada detector específico (aplicado à região em torno de cada ponto de controle equivalente dentre um conjunto de exemplos) pode não obter a nova localização da característica corretamente devido a algum ruído, mudança de iluminação ou no caso da face humana variação de expressão da facial. Uma ilustração desse tipo de problema pode ser verificada na Figura A.8.

Uma maneira simples de realizar a correspondência entre os pontos extraídos através dos detectores específicos e o modelo deformável da face, é a de projetar do vetor de pontos específicos no subespaço do modelo. Esta operação minimiza a distância entre os pontos extraídos e a forma mais plausível no subespaço.

A etapa de correspondência dos pontos deve ser repetida iterativamente até que não se encontre mudanças significativas entre a nova estimativa e anterior, isto é, até que se verifique a convergência no alinhamento entre o modelo deformável e o estimado, ou até ser atingido um número de iterações previamente estipulado.

Problema da Estimativa Inicial

Um requisito básico para que o modelo ASM consiga se ajustar à forma desejada é um bom posicionamento inicial do modelo. Acontece que, como o modelo ASM se ajusta baseado na vizinhança de cada um dos pontos de controle, caso a estimativa do posicionamento inicial seja ruim, o modelo poderá não conseguir se ajustar corretamente à forma desejada, pois o



(a) Pontos capturados pelos detectores específicos de cada ponto.

(b) Pontos após a aplicação da restrição do formato da face.

Figura A.6: Resultados antes e depois da realização da correspondência com o formato da face. Adaptado de Baggio [5].

espaço de busca do deslocamento de cada ponto do modelo poderá não ser suficientemente grande para encontrar a efetiva mudança de posicionamento. Um exemplo do que ocorre pode ser visto na figura A.7

No caso da face humana, uma abordagem bastante adotada é a de utilizar detectores genéricos de objetos para identificar o posicionamento da face para que, de posse dessa informação, o modelo ASM se ajuste melhor. Um detector de objetos genérico comumente utilizado para esta finalidade é o proposto por Viola e Jones [41].

Método de Multi-resolução

Em [74] Cootes *et al.* propõe uma alteração no algoritmo original do ASM, com a qual obtiveram melhores resultados, que consiste na utilização, tanto na fase de treinamento quanto na de ajuste do modelo, de imagens em níveis distintos de resolução criando a chamada "Pirâmide de Resolução". Assim, os perfis de intensidade são estudados tanto para imagem original quanto para níveis de resolução inferiores à original, com a finalidade dos ASM conseguirem convergir mais rápido e precisamente.

Para criar a pirâmide de resolução das imagens, aplica-se um filtro de suavização (ou borragem) Gaussiana, de dimensão $5x5$ à imagem original e reamostra a imagem resultante porém com a metade da resolução da imagem anterior, sendo repetido esse procedimento para os demais níveis da pirâmide. Sendo assim, a mudança no nível de resolução ocorre

(a) Efeito de um bom posicionamento inicial (b) Efeito de um posicionamento inicial ruim

Figura A.7: Resultados obtidos com diferentes posicionamentos iniciais do modelo. Adaptado de Cootes *et al.*[6].

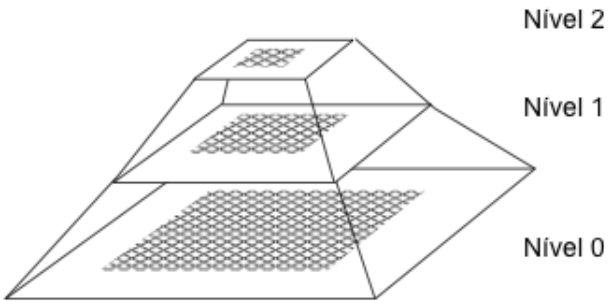


Figura A.8: Níveis de resolução. Adaptado de Cootes *et al.* [6].

quando for atingido um número máximo de iterações ou o modelo se alinhar corretamente.